Appunti del corso di Genomica

Prof.ssa Federica Gemignani

Il corso è da 6 CFU, così distribuiti: 4 CFU di lezioni frontali, 1 di esercitazioni di bioinformatica, 1 di esperienza di laboratorio.

I testi consigliati sono: "Genetica e genomica nelle scienze mediche", T. Strachan, J. Goodship, P. Chinnery, ed. Zanichelli (2016), oppure "DNA ricombinante. Geni e genomi" di J. Watson, ed. Zanichelli.

Questo, è, a grandi linee, il programma del corso:

Fondamenti del DNA. Principi e fondamenti del DNA, dei cromosomi e delle cellule. I principi che stanno alla base delle tecnologie essenziali per lo studio del DNA. Struttura fondamentale del gene, espressione genica e organizzazione del genoma umano. Le famiglie geniche e la loro origine. Gli pseudogeni processati e non processati; i retrogeni Long Interspersed NucleotideElements (LINE) e Short Interspersed Nucleotide Elements (SINE). Il genoma nucleare e il genoma mitocondriale a confronto. I principi della variabilità genetica. Le origini della variabilità della sequenza del DNA la riparazione del DNA la scala della variabilità genetica umana la variabilità genetica funzionale e il polimorfismo proteico. La straordinaria variabilità genetica e il sistema immunitario. Fondamenti della genomica. Sviluppo del progetto Genoma Umano (HGP). Cenni storici, aspetti culturali ed organizzativi Costruzione di mappe genetiche (marcatori polimorfici, STS, EST, microsatelliti) e fisiche (YAC, BAC, Cosmidi etc). Le principali strategie ed approcci scientifici utilizzati per la realizzazione dell'HGP e le sue finalità. Organizzazione di un progetto di sequenziamento su larga scala sequenziamento gerarchico e sequenziamento shotgun dell'intero genoma. Tecnologia microarray genomici e di espressione. Bioinformatica applicata alla genomica analizzare il genoma umano (Lab. informatica/lezioni interattive/esercitazioni da svolgere portando un tablet o PC in aula). Banche dati e annotazioni delle sequenze (NCBI National Center for Biotechnology Information) Database di sequenze di acidi nucleici. UCSC Genome browser. Database di sequenze proteiche. Database di malattie genetiche OMIM. Database Bibliografici. Utilizzo di database bibliografici (MEDLINE). Introduzione all'utilizzo dei Databases Genomici (Blast, DbSNP, Genome Browser). La "progettazione" dei primers per amplificare regioni del genoma. Utilizzo degli enzimi di restrizione. Dalla sequenza dei geni all'interpretazione delle sequenze. Biotecnologie applicate alla genomica umana. Gli approcci genetici al trattamento delle malattie. I principi della terapia genica ed alcuni esempi. Genetica e genomica del cancro. Il cancro caratteristiche generali e fenotipi delle cellule cancerose la genetica del cancro oncogeni e soppressori tumorali. Importanza delle biotecnologie nello sviluppo di farmaci "biologici o intelligenti". Il DNA fingerprinting e la medicina legale. Il mondo degli RNA I microRNA (e loro meccanismo di azione). MiRNA e cancro. I long non coding RNA (lncRNA). Circular RNA e il loro ruolo nel genoma.

La genomica è una branca della biologia molecolare che si occupa dello studio del genoma degli organismi viventi. In particolare si occupa della struttura, contenuto, funzione ed evoluzione del

genoma. È una scienza che si basa sulla bioinformatica per l'elaborazione e la visualizzazione dell'enorme quantità di dati che produce.

La genomica nacque negli anni 80, quando furono prese le prime iniziative per il sequenziamento di interi genomi. Una data di nascita si può probabilmente far coincidere con il sequenziamento completo del primo genoma, nel 1980: si trattava del genoma di un virus, il fago Φ-X174. Il primo sequenziamento del genoma di un organismo vero e proprio fu completato nel 1995 e si trattava di un batterio, Haemophilus influenzae, con un genoma di notevoli dimensioni (1.83 milioni di paia di basi). Da allora i genomi "completati" aumentano esponenzialmente. La prima pianta il cui genoma è completamente noto nella sua sequenza è stata Arabidopsis thaliana.

Il sequenziamento del genoma umano

Con grande rilievo fu data dalla stampa non specialistica la notizia del sequenziamento del genoma umano, nel 2001. Un progetto pubblico di sequenziamento nacque nel 1986 sotto il nome di progetto genoma umano. Nonostante l'enorme quantità di soldi spesi (il finanziamento maggiore era statunitense, pari a tre miliardi di dollari) da vari stati per i finanziamenti pubblici, questo fu pesantemente sconfitto dalla strategia innovativa di sequenziamento di un'azienda privata, la Celera Genomics, che riuscì a completare (nel 2000) il sequenziamento del genoma umano in una frazione del tempo e dei costi rispetto al progetto pubblico (300 milioni di dollari), facendo scoppiare grandi polemiche e cambiando radicalmente l'approccio seguito dai progetti pubblici. La sequenza fu pubblicata nel 2001 in un articolo su Nature, che combinava i risultati di entrambi i progetti; si trattava di una bozza pari al 90% della sequenza e ancora con notevoli probabilità di errori. Una sequenza accurata al 99,99% e pari al 99% del genoma umano è stata pubblicata nel 2003.

Obiettivi della genomica

Tra gli obiettivi che si pone la genomica vi è dunque l'allestimento di complete mappe genetiche e fisiche del DNA degli organismi viventi, proseguendo con il suo completo sequenziamento. La sequenza del DNA viene poi annotata, ovvero vengono identificati e segnalati tutti i geni e le altre porzioni di sequenza significative, insieme a tutte le informazioni conosciute su tali geni. In questo modo è possibile ritrovare in maniera organizzata ed efficace le informazioni in appositi database, normalmente accessibili via Internet gratuitamente. Grazie al sequenziamento di diversi genomi è nata la genomica comparativa, che si occupa del confronto tra i genomi di diversi organismi, nella loro organizzazione e sequenza.

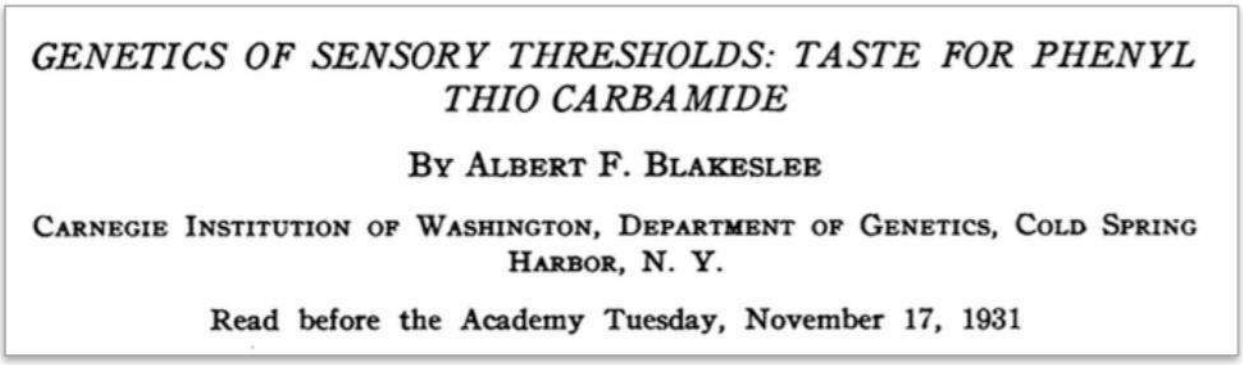
L'esperienza di laboratorio prevista dal corso, distribuita su 3 pomeriggi (estrazione DNA, PCR e gel) consisterà nella determinazione del polimorfismo del gene TAS2R38 coinvolto nella percezione dell'amaro. La percezione gustativa è determinata geneticamente. I mammiferi riconoscono soltanto 5 gusti di base: dolce, acido, amaro, salato e umami (il sapore di glutammato). Il riconoscimento dei gusti è mediato da cellule sensoriali specializzate (papille gustative) connesse direttamente al sistema nervoso centrale. La percezione del gusto è un processo che si compone di due fasi: prima, le molecole si legano a recettori specifici posizionati sulla superficie delle cellule sensoriali; poi le cellule sensoriali generano

un impulso nervoso elaborato dal cervello. Per esempio, la stimolazione di cellule sensoriali deputate a rispondere allo stimolo del dolce genera la percezione del sapore dolce nel cervello. Il gusto amaro viene riconosciuto da recettori proteici nella cui produzione sono implicati approssimativamente 30 geni differenti (più precisamente, 28 geni funzionali validati e 16 pseudogeni).

Mentre le sensazioni di dolce e salato sono convogliate attraverso canali ionici, per il dolce, l'amaro e l'umami esiste una famiglia di geni, indicati dalla sigla "TAS_R_ (TAste gene Receptors)". In particolare, i geni TAS1R codificano per i recettori di umami e dolce, mentre i TAS2R per l'amaro. Quest'ultima categoria è formata da 25 geni, che codificano per differenti sostanze amare, quali la salicina (TAS2R16), la stricnina (TAS2R10), l'aloina (TAS2R43) e la 6-nitrosaccarina (TAS2R61).

Uno dei più importanti è il gene che codifica per il recettore sensibile alla feniltiocarbammide (PTC), TAS2R38 (fra gli altri permette di discriminare e apprezzare il ricercato e particolare gusto amaro della cicoria).

La diversa sensibilità gustativa nei confronti delle sostanze dal sapore amaro (come le tiouree) è nota da tempo e risale a quando accidentalmente Fox la scoprì mentre sintetizzava nel suo laboratorio dei dolcificanti negli anni '30. Alcune polveri di PTC si dispersero nell'aria provocando delle lamentele tra i suoi colleghi a causa del sapore amaro che sentivano sulle loro labbra mentre Fox invece non avvertiva nessun sapore. Per capire meglio le differenze di sensibilità, Fox fece assaggiare la polvere di PTC a numerose altre persone (il PTC è tossico solo ad elevate concentrazioni). Rilevò che la maggior parte degli individui si divide in due categorie: quelli in grado di percepire l'amaro anche a piccole concentrazioni (tasters) e quelli incapaci di percepire l'amaro (non tasters). Gli studi di popolazioni pubblicati dal genetista Albert Blakeslee nel 1932 dimostrarono che il carattere taster è determinato geneticamente, viene ereditato come un carattere mendeliano dominante e la sua frequenza varia nelle diverse popolazioni (Blakeslee AF. 1932. Genetics of sensory thresholds: taste for phenyl thio carbamide. Proc Natl Acad Sci 18: 120-130).



GENETICS OF SENSORY THRESHOLDS: TASTE FOR PHENYL THIO CARBAMIDE

By Albert F. Blakeslee

Carnegie Institution of Washington, Department of Genetics, Cold Spring Harbor, N. Y.

Read before the Academy Tuesday, November 17, 1931

L'esame si compone di una prova scritta, della durata di 2 ore, composta da domande aperte e da esercizi di genetica o di genomica, seguita da una prova al computer, della durata di circa un'ora, che consiste in esercizi di bioinformatica. La relazione di laboratorio deve essere inviata via mail entro la data della prova scritta. La prova orale può essere sostenuta, a discrezione della docente. Se non viene ritenuta necessaria, lo studente potrà richiederla facoltativamente.

Acidi nucleici e informazione genetica

Ogni organismo vivente, unicellulare o multicellulare, è caratterizzato dalla presenza di specifiche molecole di acido nucleico (DNA o RNA) in grado di produrre una copia identica di se stesse. Esclusi alcuni virus a RNA, l'informazione genetica degli organismi viventi è contenuta nelle molecole di DNA. Il completo assetto genetico di un organismo è detto genoma. Il genoma umano è costituito da una sequenza di circa 3 miliardi di desossiribonucleotidi (le unità base del DNA), organizzati in 23 coppie di cromosomi lineari, più una piccola molecola di DNA mitocondriale circolare. Il DNA genomico deve essere duplicato prima di ogni divisione cellulare per assicurare una copia completa delle informazioni necessarie alla sintesi delle proteine e degli RNA che presiederanno al funzionamento delle cellule figlie (Box 1.1). L'essenza biologica della vita è riconducibile, in ultima analisi, alla continua replicazione e trasmissione degli acidi nucleici da una generazione all'altra di cellule.

Il dogma centrale della genetica molecolare (Box 1.1)

La sequenza dei nucleotidi del DNA contiene l'informazione necessaria alla codifica delle proteine di ogni cellula. A seconda delle necessità della cellula, il DNA dei geni viene trascritto in mRNA, che a sua volta dirige la sintesi delle corrispondenti proteine. Il cosiddetto dogma centrale della genetica molecolare è rappresentato dunque dal flusso dell'informazione che dal DNA, attraverso l'RNA, arriva alle proteine.

La replicazione del DNA, a opera della DNA polimerasi, avviene a ogni divisione cellulare per assicurare il mantenimento dell'informazione genetica in tutte le cellule; la trascrizione del DNA in RNA avviene a opera delle diverse RNA polimerasi e genera diversi tipi di RNA, codificanti e non codificanti; infine, la traduzione degli mRNA consente la sintesi proteica, che avviene sui ribosomi nel citoplasma. Il dogma centrale è stato spesso ridotto all'equazione «1 gene = 1 mRNA = 1 proteina», semplificando la complessità della sintesi degli RNA e delle proteine.

Esiste anche un processo non considerato nella formulazione originaria del dogma centrale: la retrotrascrizione dell'RNA in DNA, a opera della trascrittasi inversa, che sta alla base della moltiplicazione degli elementi trasponibili del genoma e della genesi degli pseudogeni processati.

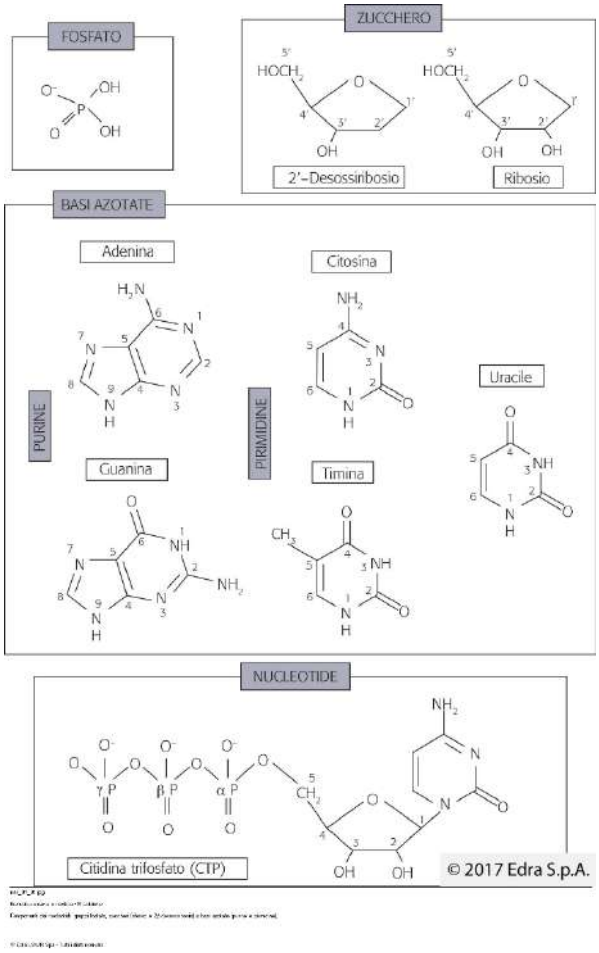
Struttura degli acidi nucleici

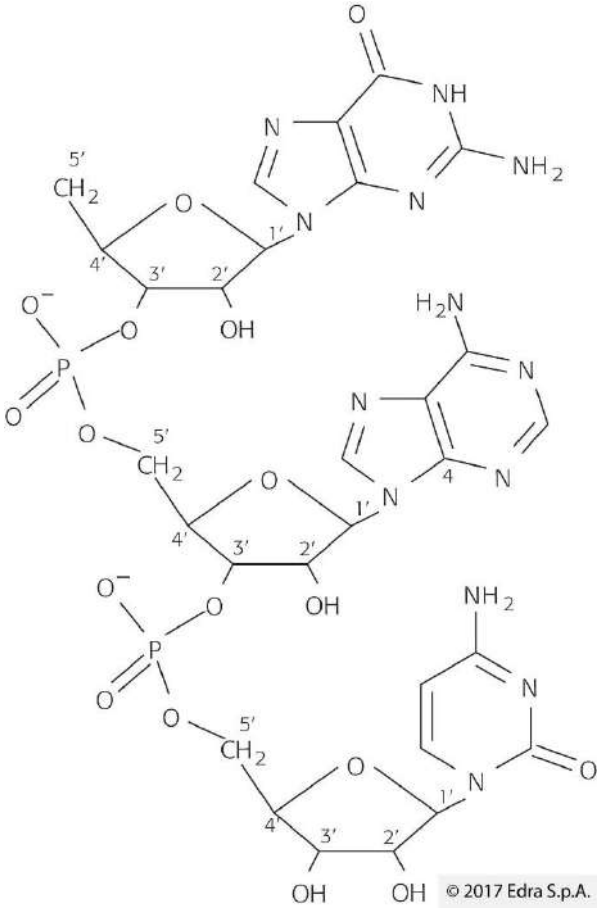
Gli acidi nucleici sono molecole polimeriche composte da unità dette nucleotidi (Figura 1.1). Nel caso dell'acido desossiribonucleico (DNA), ogni nucleotide è formato da una base azotata variabile (adenina, citosina, guanina, timina, abbreviate in A, C, G, T), unita al carbonio 1′ del 2′ -desossiribosio (uno zucchero a 5 atomi di carbonio o pentoso) a sua volta esterificato con una molecola di acido fosforico al carbonio 5'.

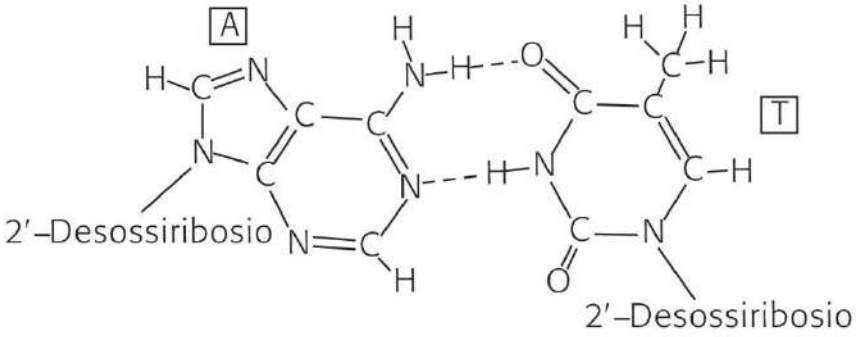
Ogni nucleotide è unito al precedente mediante il gruppo fosforico che viene esterificato anche al carbonio 3' del desossiribosio appartenente al nucleotide precedente (Figura 1.2). Nel caso dell'acido ribonucleico (RNA), lo zucchero pentoso è il ribosio (invece del 2'-desossiribosio); al posto della timina si trova una base azotata molto simile, l'uracile.

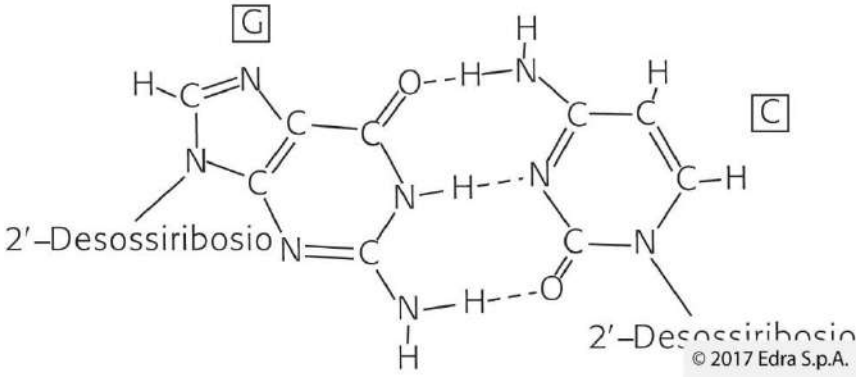
Inoltre, a differenza dell'RNA, il DNA è quasi sempre costituito da due filamenti che si avvolgono uno attorno all'altro, come in una doppia elica, e sono uniti da deboli legami elettrostatici (legami idrogeno) fra le basi azotate, dovuti all'appaiamento preciso e complementare dell'adenina con la timina e della guanina con la citosina (Figura 1.3), secondo il modello proposto da Watson e Crick nel

1953. Considerata la struttura a doppia elica del DNA, la lunghezza di una molecola di DNA è espressa in termini di paia di basi (azotate) o base pair (bp).





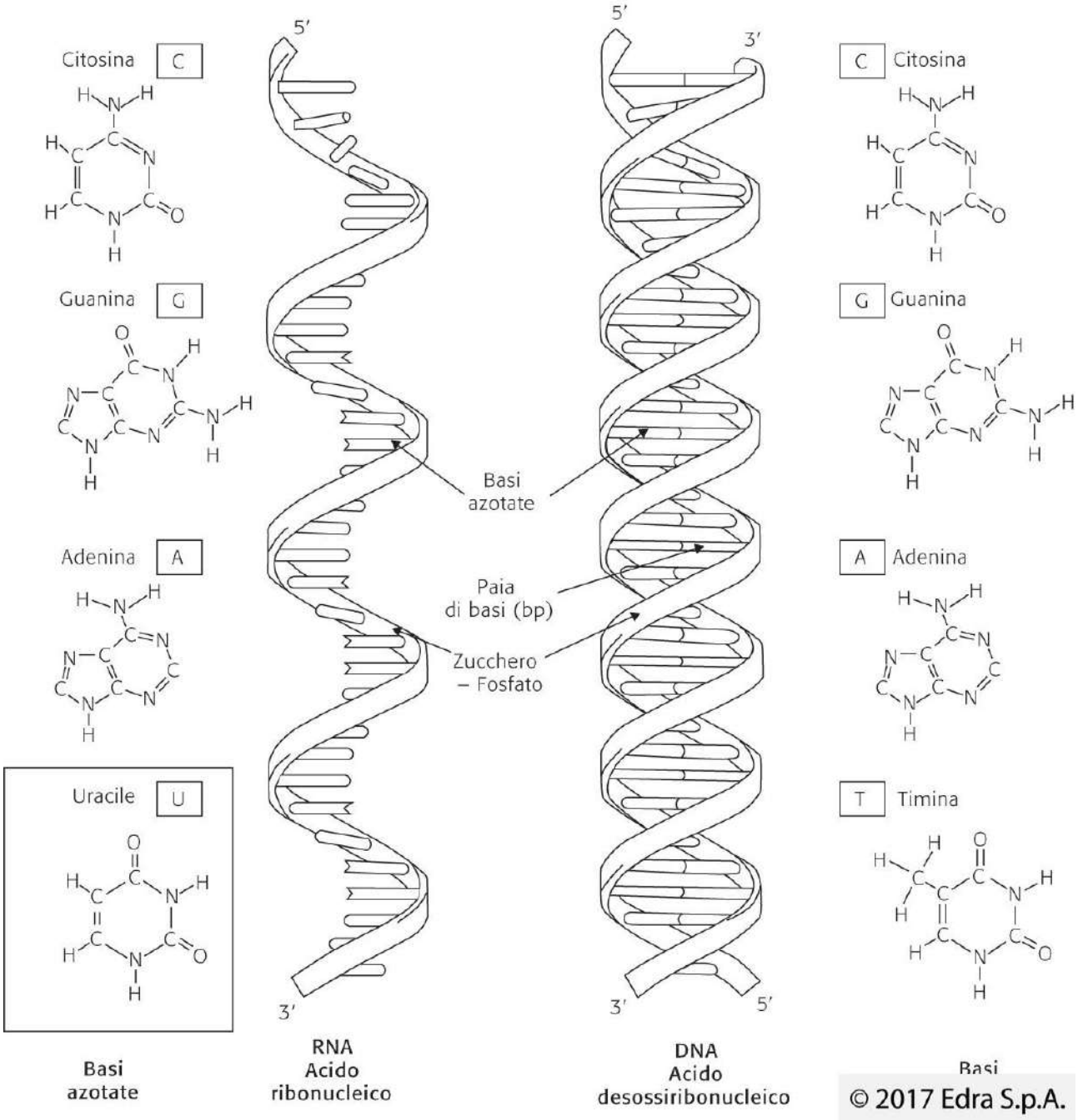




Se un frammento di DNA è dunque lungo 1.000 o 1.000.000 bp, si parlerà rispettivamente di una kilobase (kb) o di una megabase (Mb). È importante ricordare che i due filamenti del DNA sono antiparalleli, ovvero se la catena di acido fosforicodesossiribosio procede nel senso 5'-3' su di un filamento, sarà nel senso 3'-5′ sul filamento complementare (Figura 1.4).

La struttura del DNA schematizzata nella Figura 1.4 è quella della più frequente conformazione B, caratterizzata dalla formazione di un'elica destrorsa (che si avvolge in senso orario), ma in tratti ad alto contenuto di guanina e citosina (G + C) è possibile riscontrare un'elica sinistrorsa (conformazione Z).

La doppia elica, nella conformazione B, ha un diametro di circa 2 nm (10^{-9} m) e la distanza fra una coppia di basi e la successiva è di circa 3,4 Å (10^{-10} m).



© 2017 Edra S.p.A.

Il DNA all'interno delle cellule è estremamente compattato; infatti, se i 3 miliardi di nucleotidi del genoma umano fossero distesi linearmente avrebbero una lunghezza di circa 1 m. Il compattamento del DNA si realizza mediante una stretta associazione a proteine basiche dette istoni, organizzate in ottameri (costituiti da due copie degli istoni H2A, H2B, H3 e H4). Il nucleosoma rappresenta l'unità base della cromatina, ovvero l'insieme del DNA e delle proteine a esso associate, è largo 11 nm ed è costituito da 146 bp avvolte per quasi due giri intorno a un ottamero di proteine istoniche.
Un quinto istone (H1) blocca invece la doppia elica del DNA come un "fermaglio" sull'ottamero istonico.

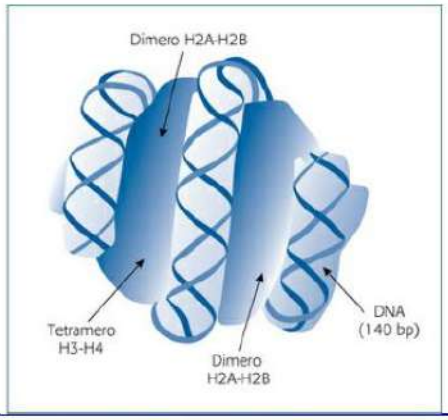


Figura 1.5 Un nucleosoma, composto da otto proteine istoniche (doppia copia di H2A, H2B, H3, H4) e circa 140 bp di DNA avvolte agli istoni con doppio giro. L'istone H1 (non mostrato) lega invece il DNA che si trova fra un nucleosoma e il successivo.

Nella maggior parte del genoma la distanza media fra un nucleosoma e l'altro di 200 bp, considerando le 146 bp avvolte al nucleosoma e circa 50 bp di DNA interposto. Al microscopio elettronico è inoltre possibile visualizzare una fibra cromatinica di circa 30 nm di spessore (detta fibra "a solenoide"), costituita da 6 nucleosomi che si succedono in modo elicoidale per ogni giro. Questi due gradi di avvolgimento del DNA consentono di ridurre di circa 40-50 volte la lunghezza totale del genoma, riducendola a 2-3 cm. E' evidente che sono necessari ulteriori gradi di compattamento della cromatina per consentire all'intero genoma di trovare spazio all'interno dei pochi micrometri di diametro del nucleo cellulare. Infatti, fra una divisione cellulare e l'altra (interfase), la cromatina appare come un gomitolo più o meno compatto. Si distinguono due tipi di cromatina: l'eterocromatina, che rimane altamente condensata durante la maggior parte del ciclo cellulare ed è trascrizionalmente inattiva, e l'eucromatina, che è meno compatta e consente sia la trascrizione del DNA in RNA sia la duplicazione del DNA durante la fase S. Piccole modificazioni covalenti degli istoni, come l'aggiunta di gruppi metile, acetile o fosfato, correlano con lo stato più o meno attivo della cromatina. Durante la divisione cellulare (mitosi), il grado di compattamento della cromatina è massimo e i singoli cromosomi si rendono finalmente visibili.

Replicazione del DNA

La complementarità dei filamenti di DNA suggerisce anche il meccanismo di sintesi (replicazione). Gli stessi Watson e Crick hanno fin dal 1953 ipotizzato che ciascun filamento di DNA funzioni da stampo per la sintesi di filamenti complementari. Da una doppia elica del DNA si generano due doppie eliche, ciascuna contenente un filamento che apparteneva alla molecola parentale e un filamento neosintetizzato. Per questo motivo il processo replicativo è detto semiconservativo.
La replicazione del DNA inizia in punti specifici, detti origini di replicazione, nei quali i due filamenti della molecola di DNA parentale vengono separati da enzimi, detti elicasi, formando una bolla di replicazione. L'enzima DNA polimerasi catalizza la sintesi dei nuovi filamenti di DNA utilizzando come precursori nucleotidici i quattro desossinucleotidi trifosfati (dATP, dCTP, dGTP, dTTP). Poiché i due

filamenti del DNA parentale sono antiparalleli, ne consegue che i due filamenti neosintetizzati debbano correre in direzione opposte.

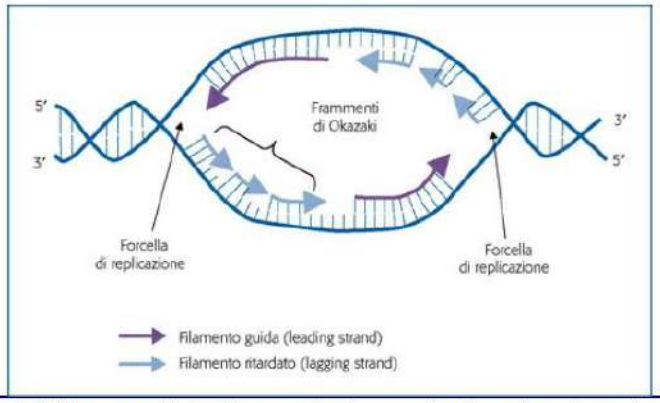
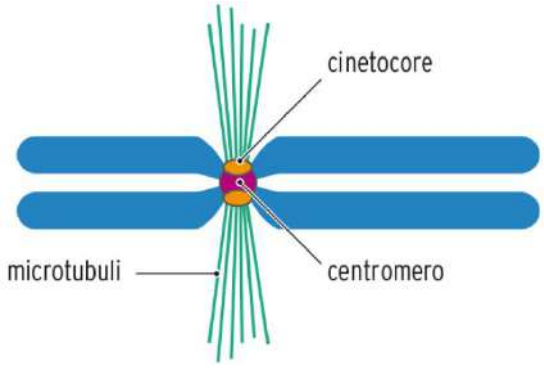


Figura 1.6 Sintesi del DNA: il filamento guida (leading strand) può essere sintetizzato in modo continuo dato che la doppia elica del DNA viene svolta davanti a esso dalle elicasi (freccia viola). Il filamento ritardato (lagging strand) deve essere sintetizzato in modo discontinuo (sempre in direzione 5'-3') a partire da piccoli primer di RNA (freccia nera) ed è composto da frammenti detti di Okazaki, successivamente uniti dalla DNA ligasi.

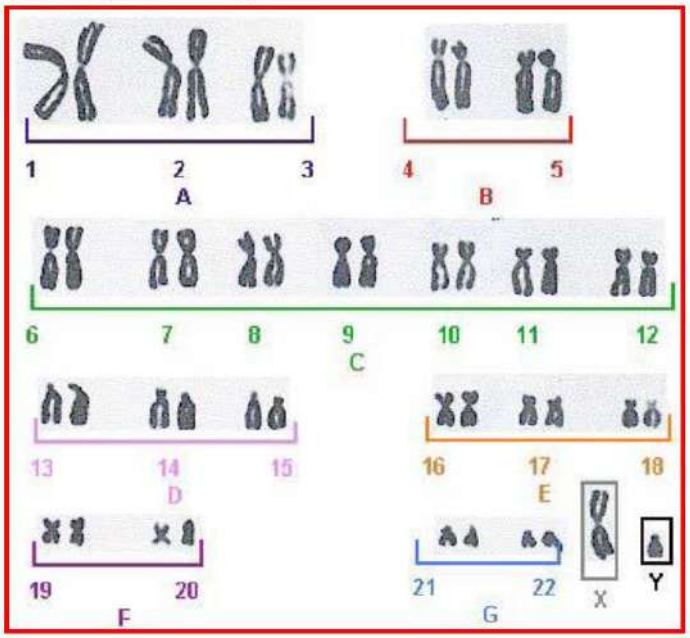
Poiché la DNA polimerasi catalizza l'allungamento delle catene esclusivamente in direzione 5'-3', su un filamento, detto filamento veloce (leading strand), la sintesi può procedere direttamente in questo senso, mentre nell'altro, detto filamento lento (lagging strand), la sintesi 5'-3' procede in direzione opposta a quella in cui si muove la forcella di replicazione. La sintesi del filamento lento avviene sotto forma di una serie progressiva di piccoli frammenti di circa 100-1.000 nucleotidi, detti frammenti di Okazaki, che vengono successivamente uniti per mezzo della DNA ligasi. Poiché solamente il filamento veloce viene sintetizzato in modo continuo, si dice che la sintesi dei filamenti di DNA è semidiscontinua. Dai numerosi punti di origine la replicazione procede in maniera bidirezionale, formando bolle di replicazione che infine si fondono. La completa replicazione del DNA delle cellule umane in coltura richiede in genere circa 8 ore. Le estremità di cromosomi di DNA (telomeri) richiedono un meccanismo diverso di replicazione perché manca lo stampo per l'allungamento del filamento lento. La struttura dei telomeri eucariotici è particolare in quanto consiste in una lunga sequenza di unità ripetute. Nell'uomo la sequenza base di tali unità è TTAGGG. Queste ripetizioni, oltre a permettere la replicazione, hanno anche la funzione di proteggere le estremità dei cromosomi dalla degradazione. Un enzima specifico, la telomerasi, utilizzando un RNA stampo, estende il filamento veloce, che a sua volta funziona successivamente da stampo per la sintesi del filamento lento. La telomerasi è particolarmente attiva nelle cellule germinali, ma lo è molto meno nella maggior parte dei tessuti somatici, determinando così la diminuzione della lunghezza dei telomeri con i cicli successivi di replicazione.

La struttura dei cromosomi



I cromosomi umani esaminati in mitosi sono classificati e ordinati in base alla lunghezza e alla posizione del centromero, in accordo con la classificazione di Denver, proposta nel 1960 al Congresso di Genetica Umana (Colorado).

I cromosomi sono stati suddivisi in 7 gruppi in ordine decrescente di lunghezza.



Cariogramma di cromosomi metafasici umani ordinati secondo la classificazione di Denver.

Nella metafase i cromosomi appaiono costituiti da due subunità longitudinali identiche, o cromatidi fratelli, unite a livello del centromero costituite ciascuna da una doppia elica di DNA.

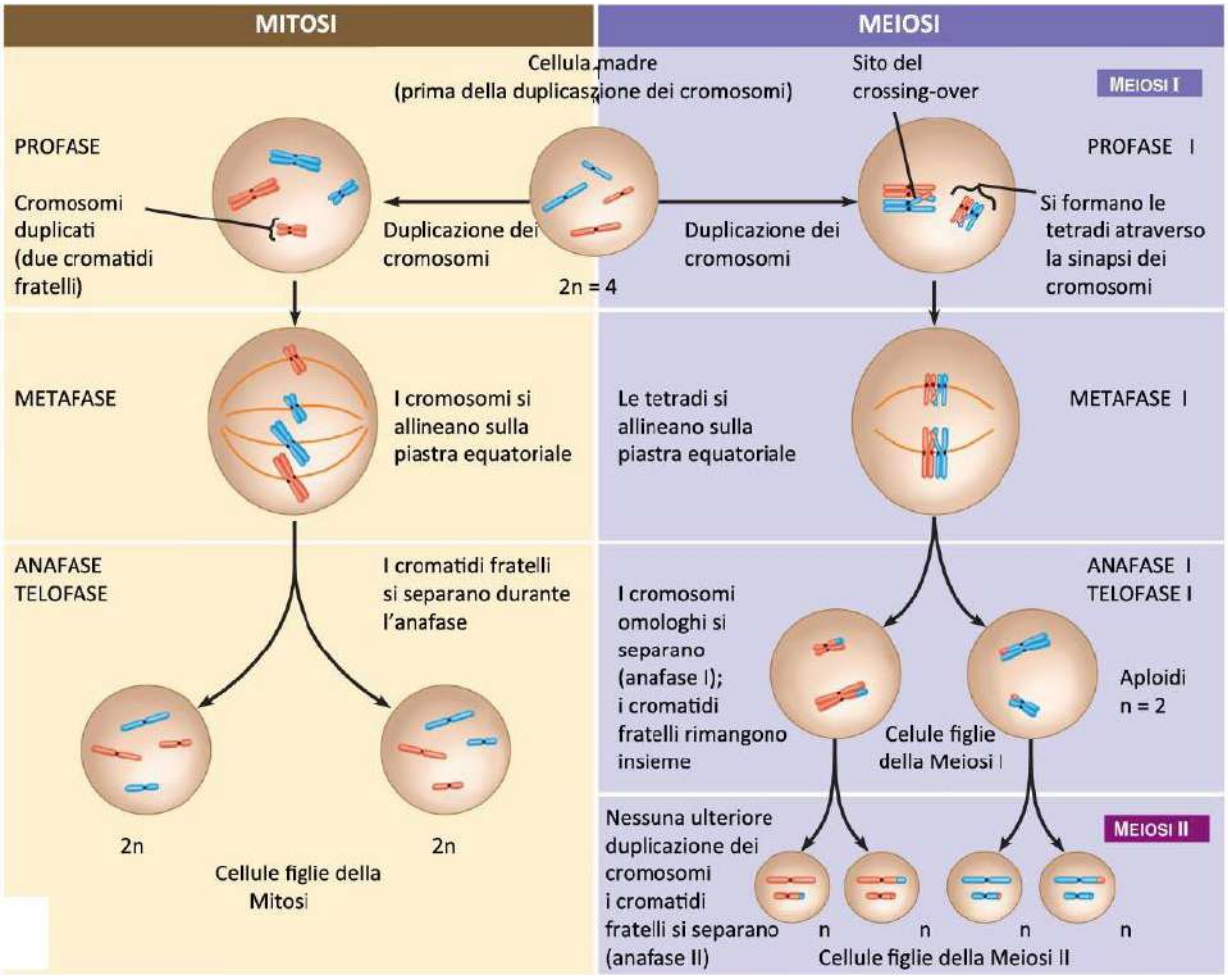
La posizione relativa del centromero è costante e divide ciascun cromatidio in due bracci:

p (da *petit*, corto)

e

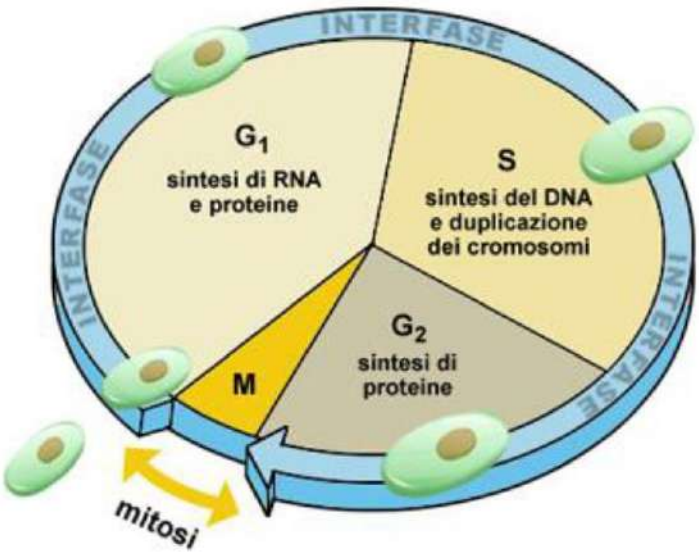
q (da *queue*, lungo).

Mitosi e meiosi

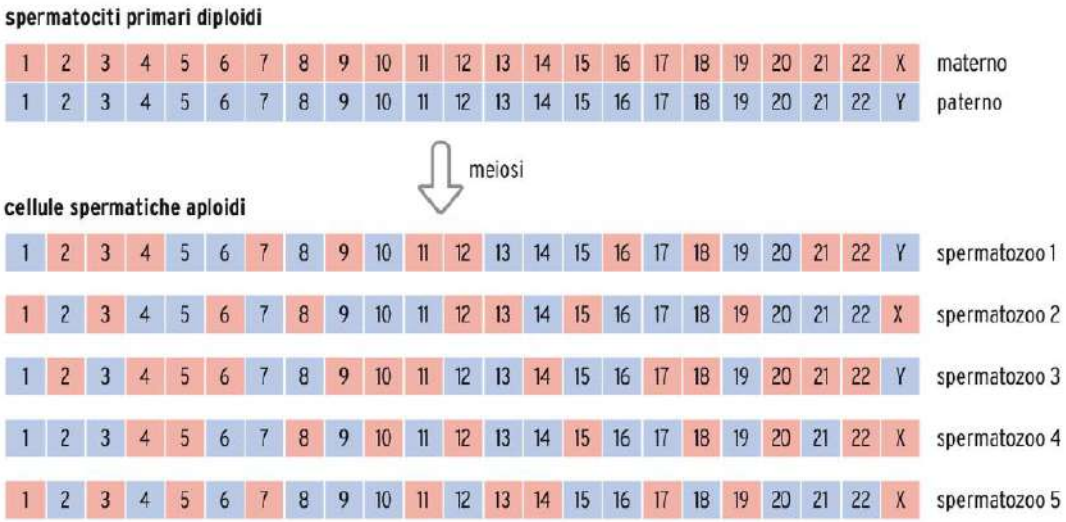


I due processi di divisione cellulare, nelle cellule somatiche e in quelle sessuali (spermatozoi e cellule uovo).

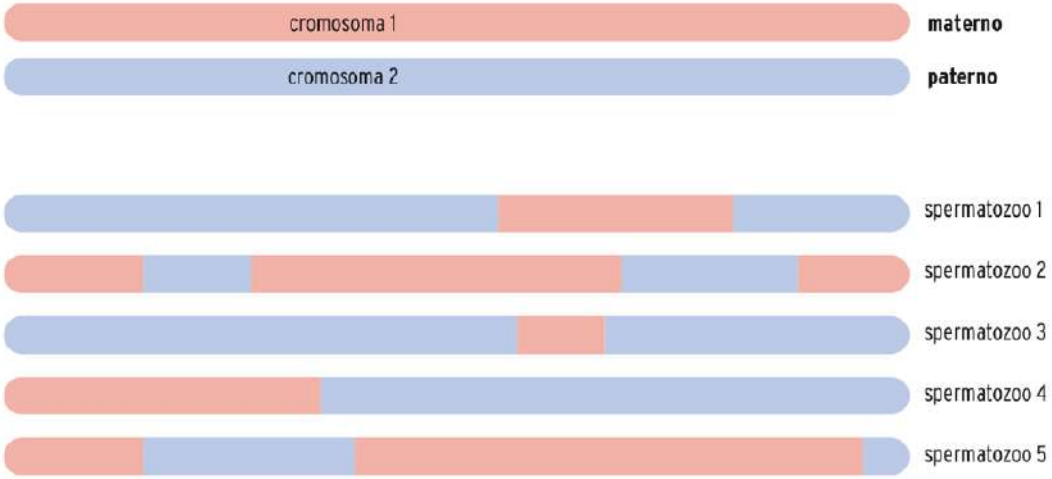
Il ciclo cellulare



Ogni nostro gamete è unico per l'assortimento indipendente degli omologhi...



... e la ricombinazione.



La bioinformatica

Queste esercitazioni hanno lo scopo di fornire gli strumenti ed i metodi per accedere all'informazione biologica in modo razionale ed efficiente, utilizzando le risorse disponibili in rete.

La bioinformatica è una disciplina che fa da ponte tra la biologia e l'informatica. E' l'applicazione dell'informatica alla gestione e all'analisi dei dati biologici.

Lo sviluppo della bioinformatica è stato determinato da vari fattori:

1. Lo sviluppo di biotecnologie innovative, con il sequenziamento automatico del DNA e nuovi approcci di tipo "genomico" e "proteomico" con la miniaturizzazione dei saggi biologici e l'uso di tecniche quali microarray;
2. Lo sviluppo di potenzialità informatiche (hardware), grazie a una maggiore potenza e velocità di elaborazione dei dati dei computer e all'ampliamento e potenziamento delle reti di telecomunicazioni;
3. Lo sviluppo di nuovi programmi informatici (software).

I database (o banche dati) sono archivi digitali in cui valgono le stesse proprietà degli schedari. È importante avere un contrassegno che identifica in modo univoco la scheda che contiene i dati. Questo "contrassegno" (chiave) spesso viene chiamato "ID" e in alcuni database biologici viene chiamato "AC" (Accession number). Un esempio è sul sito dell'NCBI: www.ncbi.nlm.nih.gov .

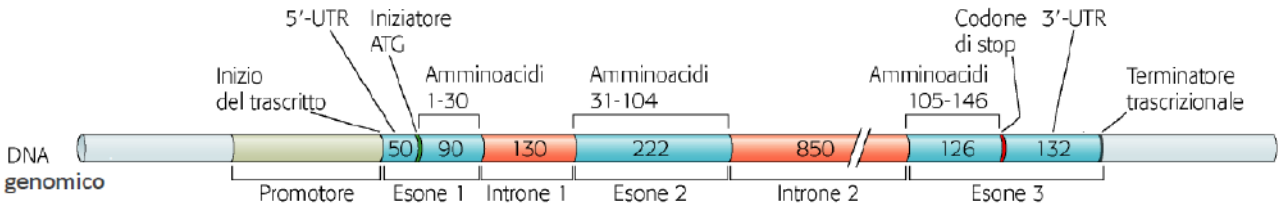
Pubmed è un database di citazioni bibliografiche. Le singole citazioni bibliografiche sono organizzate in campi (Autore,titolo,rivista...). Lo scopo principale di un database è la gestione dell'informazione.

Struttura ed espressione dei geni

Dal punto di vista della genetica molecolare, per gene si intende una sequenza di DNA potenzialmente trascrivibile in un RNA funzionalmente attivo. Tale RNA può svolgere direttamente una funzione strutturale e/o catalitica (rRNA, tRNA, ecc.) oppure trasportare l'informazione per la sintesi di una proteina (mRNA). Nel genoma umano si stiama che siano presenti circa 20 00 geni codificanti proteine e almeno altrettanti geni che vengono trascritti in RNA non codificanti (ncRNA), che svolgono un ruolo importante nella regolazione della conformazione della cromatina e della trascrizione degli stessi geni codificanti proteine. Il genoma umano si divide tra genoma nucleare e genoma mitocondriale. Il genoma nucleare è composto da sequenze altamente conservate e sequenze poco conservate. Tra le prime si annoverano i geni codificanti proteine (1,1% del totale del genoma).

Studi hanno dimostrato che, mediamente, i geni umani contengono introni più lunghi dei geni di C. elegans e D. melanogaster. Inoltre, la porzione coding, nell'uomo, è molto più ridotta rispetto a quella non coding, l'opposto di ciò che avviene negli organismi più primitivi, come microrganismi e Invertebrati.

Osserviamo la struttura tipica di un gene, prendendo ad esempio l'organizzazione del gene della β-globina.



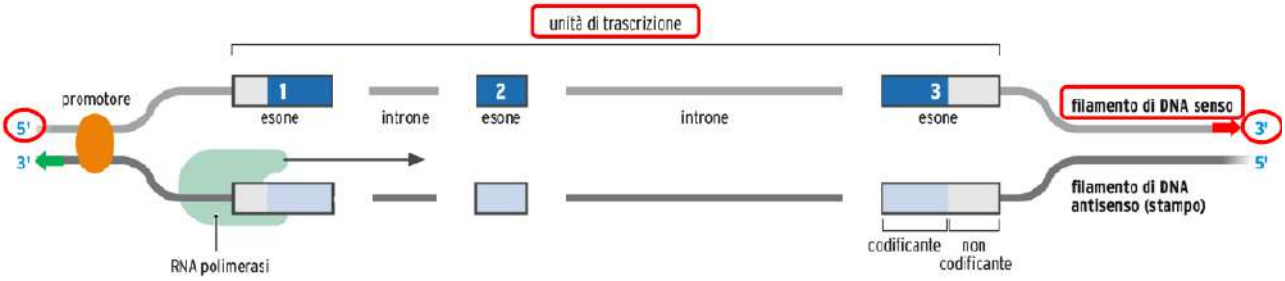
Il filamento in figura si estende dalla 5' alla 3'. La regione a monte del sito inizio della trascrizione è detto promotore. La numerazione dei nucleotidi del promotore inizia da -1, che corrisponde al nucleotide che precede il sito di inizio della trascrizione (indicato con +1). In questa regione, di lunghezza variabile, si trova una serie di brevi sequenze (elementi cis) che viene riconosciuta e legata da fattori di trascrizione (elementi trans). I fattori di trascrizione reclutano l'RNA polimerasi II al sito giusto per iniziare la trascrizione grazie all'interazione del mediator complex, un complesso multiproteico contenente circa 30 proteine. Tale complesso è assolutamente necessario per l'inizio della sintesi dell'RNA, ma anche per le fasi successive della trascrizione. I geni che presentano elevati livelli di trascrizione, come gli istoni o la β-globina, hanno elementi del promotore che includono sempre un TATA box, spesso TATAAA o una sua variante, a circa -25 bp dal sito di inizio della trascrizione. I promotori di molti altri geni, tra cui i cosiddetti geni housekeeping, ossia geni necessari al funzionamento generale della cellula, non hanno TATA box, ma presentano spesso altri elementi come i GC box (GGGCGG). Tra gli elementi frequentemente presenti nel promotore ricordiamo il CAAT box (circa -80 bp dal sito di inizio della trascrizione).

La sequenza codificante della maggior parte dei geni umani è suddivisa in segmenti detti esoni, separati da sequenze interposte non codificanti, detti introni. Il primo esone comincia al sito di inizio della trascrizione, ma il primo tratto (lungo in genere 200-300 bp) non è codificante; questo segmento pertanto è trascritto, ma non tradotto e viene indicato come regione 5'UTR (UnTranslated Region). La regione 5' UTR è importante per l'efficienza della traduzione in quanto facilita il legame dell'mRNA ai ribosomi. La regione tradotta inizia di solito con il trinucleotide ATG nel primo esone. Il numero degli esoni presenti nei geni umani è altamente variabile. Vi sono geni piccoli costituiti da un singolo esone, come i geni per gli istoni, e altri che possiedono più di 100 esoni, come, per esempio, alcuni geni che codificano per le catene del collagene. Il numero medio è circa 9-10 esoni per gene. I singoli esoni sono generalmente piuttosto piccoli, con dimensioni medie di circa 200 bp, ma esistono alcuni esoni ecccezionalmente lunghi, che possono superare anche le 5 kb. Al contrario degli esoni, la dimensione degli introni è molto più variabile. In genere, i geni piccoli hanno introni piccoli, mentre in quelli più grandi gli introni possono avere una lunghezza anche di 10-20 kb. L'ultimo esone, così come il primo, contiene una sequenza trascritta, ma non tradotta, detta regione 3' UTR (generalmente più lunga della 5' UTR). La regione 3'UTR contiene il segnale di poliadenilazione (AATAAA) e sequenze dalle quali dipende la stabilità delle molecole di mRNA.

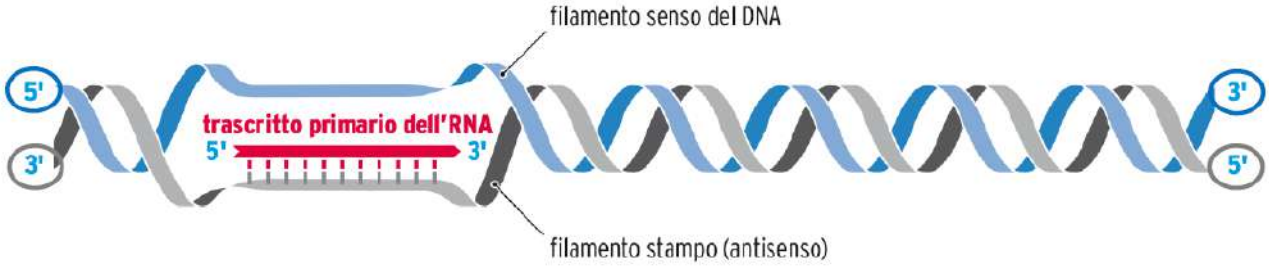
Trascrizione e maturazione degli RNA

L'informazione genetica contenuta nelle sequenze del DNA viene trasferita all'RNA e dall'RNA al polipeptide (proteina) corrispondente. Il flusso dell'informazione genetica è pertanto unidirezionale (DNA -> RNA -> proteine). Questo processo è anche noto come il dogma centrale della biologia molecolare.

Il trasferimento dell'informazione genetica dal DNA all'RNA è detto trascrizione. Durante questo processo un complesso proteico, comprendente l'RNA polimerasi, sintetizza le molecole di RNA sullo



stampo delle sequenze di DNA che costituiscono le unità di trascrizione. La RNA polimerasi si lega al sito di inizio della trascrizione richiamata da altre proteine, dette fattori di trascrizione. Questi fattori, mediante l'interazione con brevi sequenze di DNA presenti nella regione a monte dell'inizio della trascrizione (promotore), servono a posizionare la RNA polimerasi nel sito giusto e a separare i due filamenti del DNA per formare la bolla di trascrizione.
L'enzima usa come stampo uno dei due filamenti di DNA (filamento antisenso) e sintetizza una molecola di RNA in direzione 5'-3', catalizzano il legame fosfodiestere tra il gruppo ossidrilico legato al carbonio 3' del ribonucleotide precedente e il fosfato del nuovo ribonucleotide.



Il processo continua fino a che la polimerasi incontra una sequenza di arresto. A questo punto, si stacca e libera la catena di RNA, mentre la bolla di trascrizione si richiude e il DNA riassume la propria conformazione a doppia elica. L'RNA neosintetizzato ha la sequenza di basi identica a quello di uno dei due filamenti di DNA (il filamento senso), anche se la timina è sostituita dall'uracile. Da uno stesso gene possono essere trascritte consecutivamente numerose copie di RNA e il livello di trascrizione dipende da complessi meccanismi.

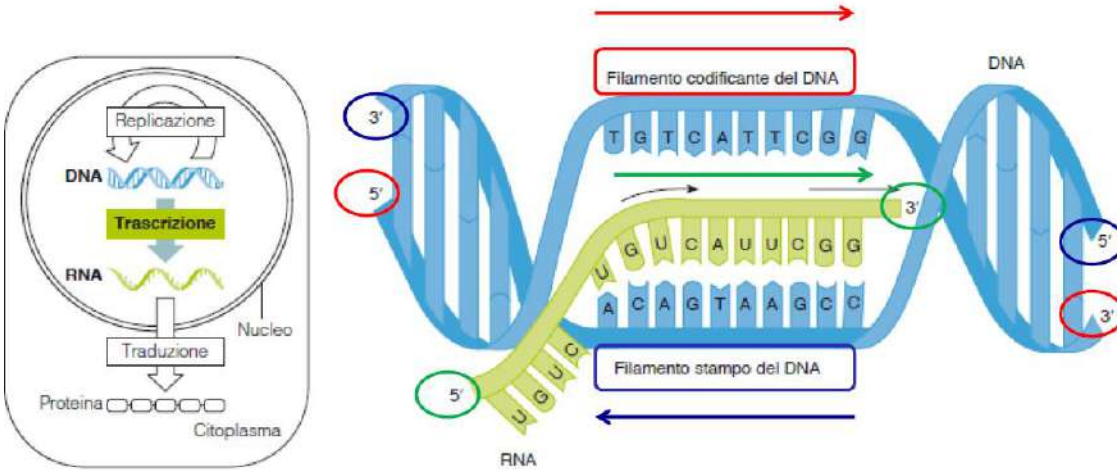


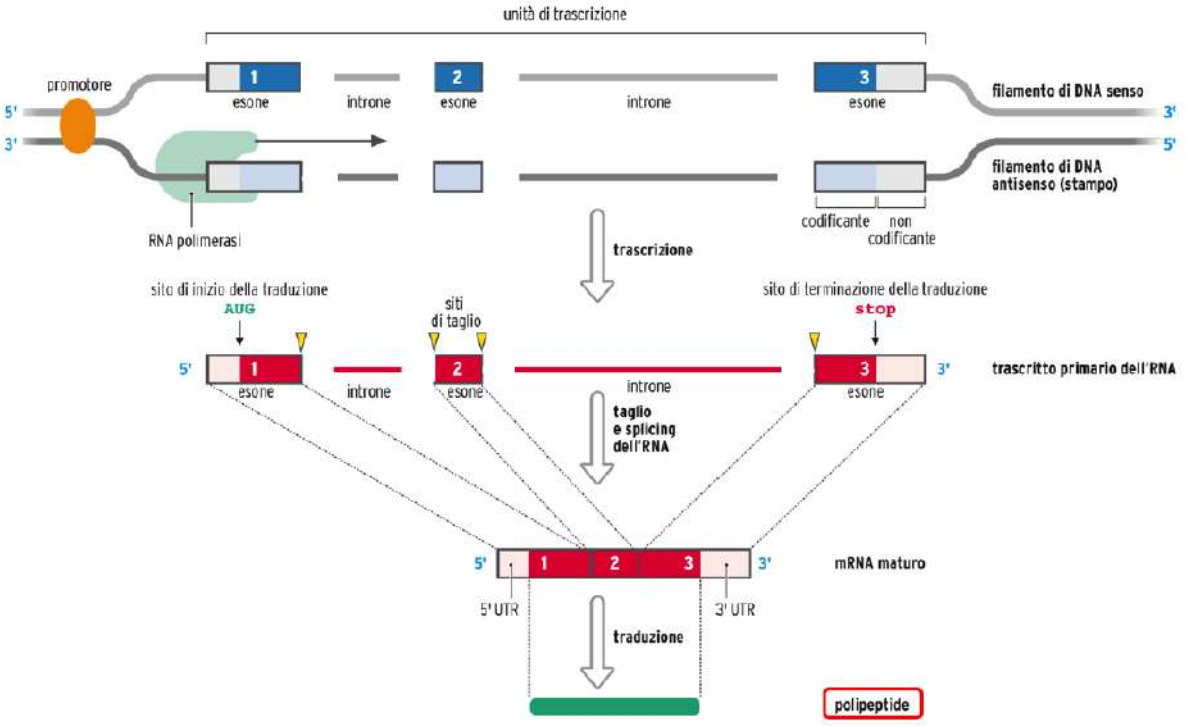
Figura 10.2 **Il rapporto tra RNA, filamento stampo del DNA e filamento codificante del DNA.** La sequenza di RNA è complementare al filamento stampo del DNA. Questa è la stessa sequenza del filamento codificante del DNA, con l'uracile (U) al posto della timina (T).

E' importante ricordare che le cellule eucariotiche possiedono tre tipi di RNA polimerasi: la RNA polimerasi I trascrive i geni degli RNA ribosomiali 18S, 5.8S e 28S (dove S sta per Svedberg ed è una misura di della sedimentazione in ultracentrifuga proporzionale alla massa); la RNA polimerasi II trascrive i geni che codificano proteine sintetizzando i precursori degli RNA messaggeri (mRNA) e anche alcuni piccoli RNA; la RNA polimerasi III trascrive i geni di tutti gli RNA transfer (tRNA), l'RNA ribosomiale 5S e altri piccoli RNA.

I precursori degli mRNA neosintetizzati dalla RNA polimerasi II (trascritti primari) devono subire una serie di modificazioni prima di essere trasferiti nel citoplasma per venire tradotti sui ribosomi. Questo processo di maturazione degli mRNA include le seguenti modificazioni.

- ✓ Aggiunta all'estremità 5' di un cap o cappuccio. Al primo nucleotide all'estremità 5' della molecola di mRNA nascente viene rimosso il fosfato terminale e viene aggiunta una molecola di guanosina monofosfato (GMP) metilata in posizione 7'. Il capping serve per proteggere il trascritto dall'attacco delle esonucleasi, che lo degraderebbero, e per facilitare il trasporto dal nucleo al citoplasma.
- ✓ Rimozione di alcune sequenze che non vengono tradotte (processo di splicing).
- ✓ Aggiunta all'estremità 3' di una coda di poli-A. La maggior parte delle unità di trascrizione ha una breve sequenza (AATAAA) che specifica il termine della trascrizione. Circa 15-30 nucleotidi a valle di questo sito, l'RNA neosintetizzato viene scisso da un enzima, un'endonucleasi, e alla molecola di RNA vengono aggiunti circa 200 residui di adenosina monofosfato (AMP). La coda poli-A ha lo scopo di stabilizzare le molecole degli mRNA maturi e di facilitare il loro trasporto dal nucleo al citoplasma. Gli mRNA codificanti gli istoni "canonici" (ovvero H2A, H2B, H3, H4 e H1) fanno eccezione a questa regola e, pur essendo trascritti dalla RNA polimerasi II, sono privi della coda poli-A.

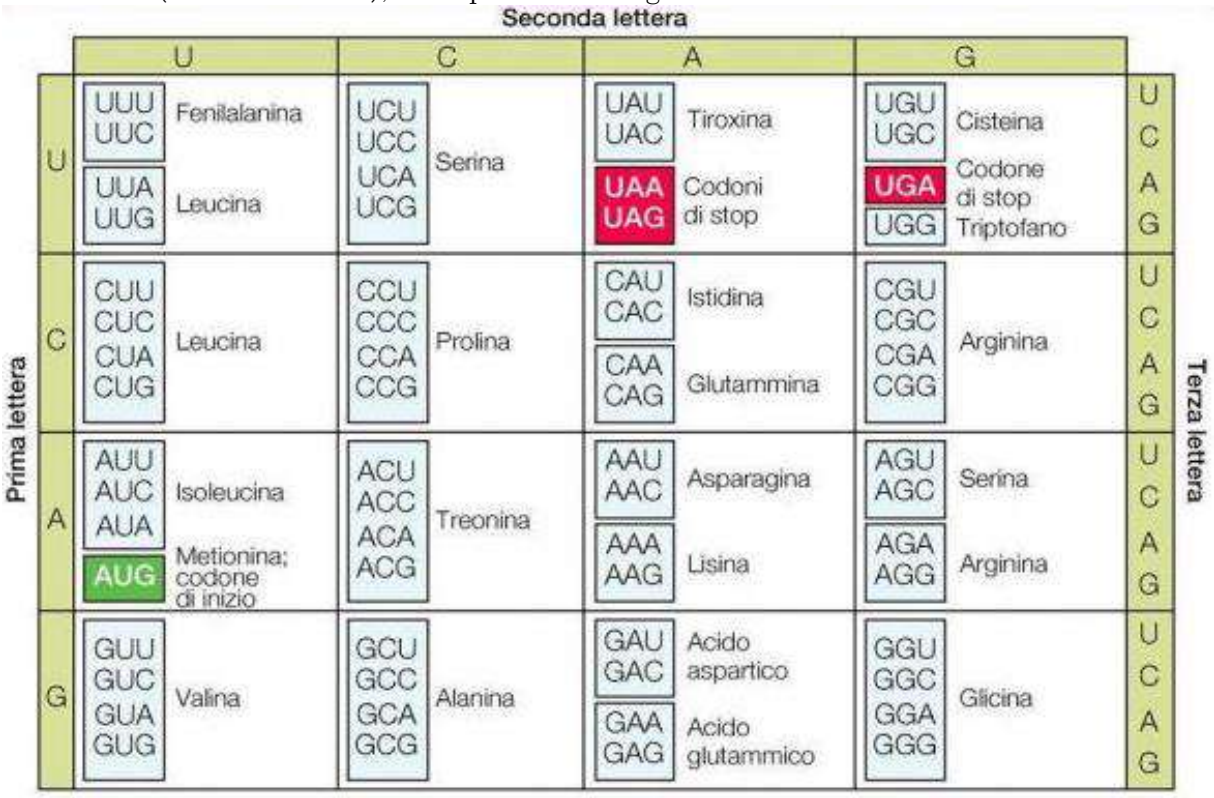
Altre importanti classi di RNA non vengono tradotte in proteine. Tali RNA, fondamentali per la sopravvivenza della cellula, includono: gli RNA ribosomiali (rRNA), gli RNA transfer (tRNA), i piccoli RNA nucleari (snRNA) e citoplasmatici (scRNA) e altre molecole di RNA non codificante (ncRNA).



Sintesi delle proteine e codice genetico

Gli mRNA maturi migrano nel citoplasma, dove dirigono la sintesi delle proteine. Essi si legano ai ribosomi, costituiti da due complessi ribonucleoproteici, detti subunità ribosomiali. La subunità grande, detta 60S, contiene circa 50 proteine e tre tipi di rRNA: 28S, 5.8S e 5S. La subunità piccola, detta 40S, contiene l'rRNA 18S e circa 30 proteine.

La sequenza amminoacidica della proteina nascente è determinata dalla sequenza nucleotidica dell'mRNA grazie a un codice genetico universale basato su triplette, ovvero gruppi di tre nucleotidi, detti codoni (unità codificanti), corrispondenti a singoli amminoacidi.



Prima lettera	Seconda lettera: U	Seconda lettera: C	Seconda lettera: A	Seconda lettera: G	Terza lettera
U	UUU, UUC: Fenilalanina	UCU, UCC, UCA, UCG: Serina	UAU, UAC: Tiroxina	UGU, UGC: Cisteina	U C A G
	UUA, UUG: Leucina		UAA, UAG: Codoni di stop	UGA: Codone di stop; UGG: Triptofano	
C	CUU, CUC, CUA, CUG: Leucina	CCU, CCC, CCA, CCG: Prolina	CAU, CAC: Istidina	CGU, CGC, CGA, CGG: Arginina	U C A G
			CAA, CAG: Glutammina		
A	AUU, AUC, AUA: Isoleucina	ACU, ACC, ACA, ACG: Treonina	AAU, AAC: Asparagina	AGU, AGC: Serina	U C A G
	AUG: Metionina; codone di inizio		AAA, AAG: Lisina	AGA, AGG: Arginina	
G	GUU, GUC, GUA, GUG: Valina	GCU, GCC, GCA, GCG: Alanina	GAU, GAC: Acido aspartico	GGU, GGC, GGA, GGG: Glicina	U C A G
			GAA, GAG: Acido glutammico		

Le molecole di tRNA veicolano gli amminoacidi ai ribosomi; ciascuna molecola di tRNA ha una struttura simile a un trifoglio e lega con l'estremità 3' uno specifico amminoacido, mentre sul versante opposto presenta una specifica tripletta di basi, detta anticodone. L'anticodone è complementare e, quindi, si lega alla tripletta del codone presente sull'mRNA. La sintesi del polipeptide ha inizio con l'interazione dell'anticodone del primo tRNA, al quale è legata la metionina, con il codone d'inizio (AUG) sull'mRNA. In particolari condizioni però è stata osservata la traduzione di peptidi che non utilizzano il codone di start canonico (AUG), come nel caso della Repeat-Associated Non-AUG mediated translation. L'amminoacido successivo viene aggiunto al primo mediante un legame peptidico tra il gruppo carbossilico del primo amminoacido e il gruppo ammino del successivo, la cui formazione è catalizzata dalla subunità grande del ribosoma che agisce come ribozima.

Vengono, quindi, incorporati nella catena polipeptidica nascente tutti gli amminoacidi, finchè si incontra un codone di termine (codone di stop). Vi sono 64 codoni possibili (in ogni posizione della tripletta c'è la possibilità di 4 basi diverse). Dei 64 codoni, 61 specificano i 20 diversi amminoacidi e 3 sono codoni di stop (UAA, UAG, UGA).

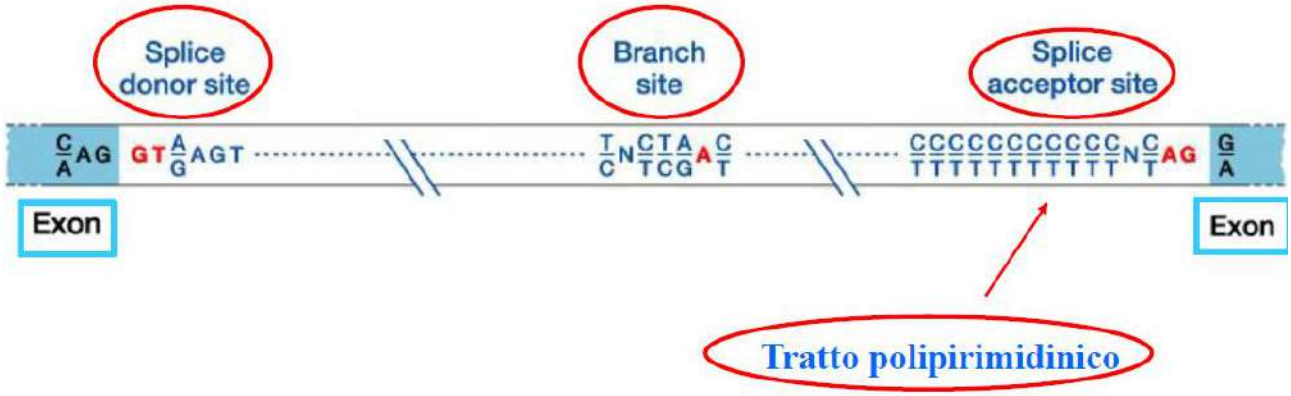
Poiché i singoli amminoacidi possono essere specificati da più codoni, si dice che il codice genetico è degenerato. Anche il numero di molecole di tRNA è inferiore a 61 (sono circa 30), ma tutti i 61 codoni possono essere riconosciuti perché l'anitcodone del tRNA si appaia specificamente soltanto con le prime due basi del codone sull'mRNA, mentre la terza base può legarsi anche ad altre basi.

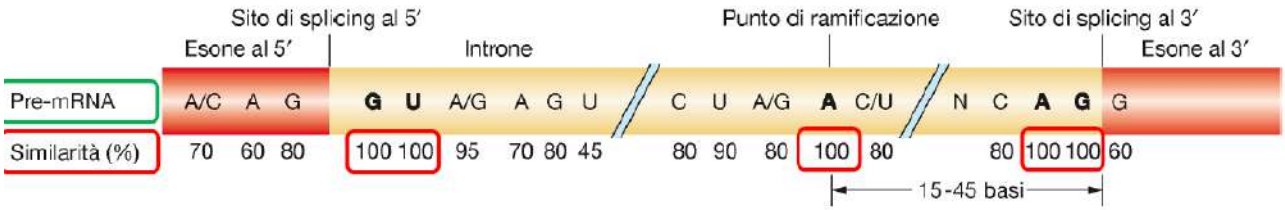
Lo splicing

Il processo di splicing consiste nella rimozione degli introni dai trascritti primari.

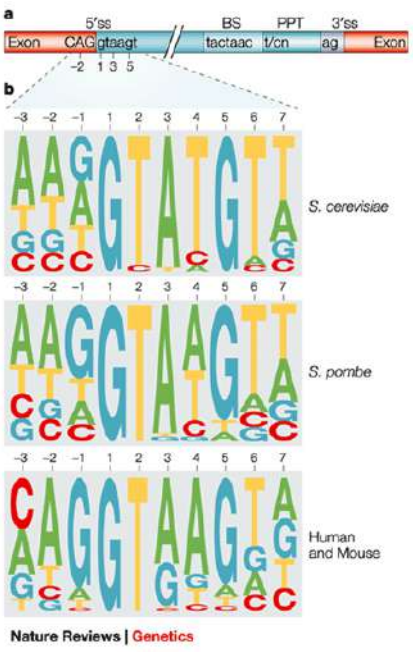
Quasi tutti gli introni cominciano con il dinucleotide GT (sito donatore di splicing) e terminano con AG (sito accettore di splicing), secondo la regola GT-AG. Questi dinucleotidi sono circondati da sequenze

consenso (come ad esempio il tratto polipirimidinico, ricco di C o di T, subito a monte del sito accettore), altamente conservate nell'evoluzione e molto simili tra loro. Questi motivi "conservati" determinano la forza del sito di splicing. Una terza sequenza consenso si trova nell'introne circa 40 bp a monte del dinucleotide AG. Essa è detta sito di ramificazione (branch site) e contiene obbligatoriamente una adenina. Tutte e tre le sequenze consenso vengono riconosciute dai fattori di splicing e sono indispensabili per una corretta eliminazione delle sequenze introniche durante la maturazione degli mRNA.





I segnali dei siti di splicing sono ben conservati tra le specie. Probabilmente la comparsa del meccanismo di splicing è molto antica.



Il processo di splicing deve essere molto preciso, dato che lo spostamento anche di un singolo nucleotide determinerebbe lo slittamento del modulo di lettura e, quindi, la sintesi di una proteina alterata.

Nel processo di splicing si verifica prima il taglio al 5' dell'introne (sito donatore), poi l'estremità libera si ripiega su se stessa formando una struttura simile a un laccio e legandosi all'interno dell'introne (sito di branching), quindi avviene il taglio al 3' dell'introne (sito accettore), e i due esoni si uniscono mentre l'introne va perso. La struttura macromolecolare che promuove e controlla le reazioni dello splicing è lo spliceosoma, formato da più

di 150 proteine e 5 RNA. Questo complesso comprende varie subunità di piccole particelle ribonucleoproteiche (snRNP), ciascuna costituita da una o due molecole di piccoli RNA nucleari, ricche di uridine (U snRNA) e da una serie di proteine specifiche.

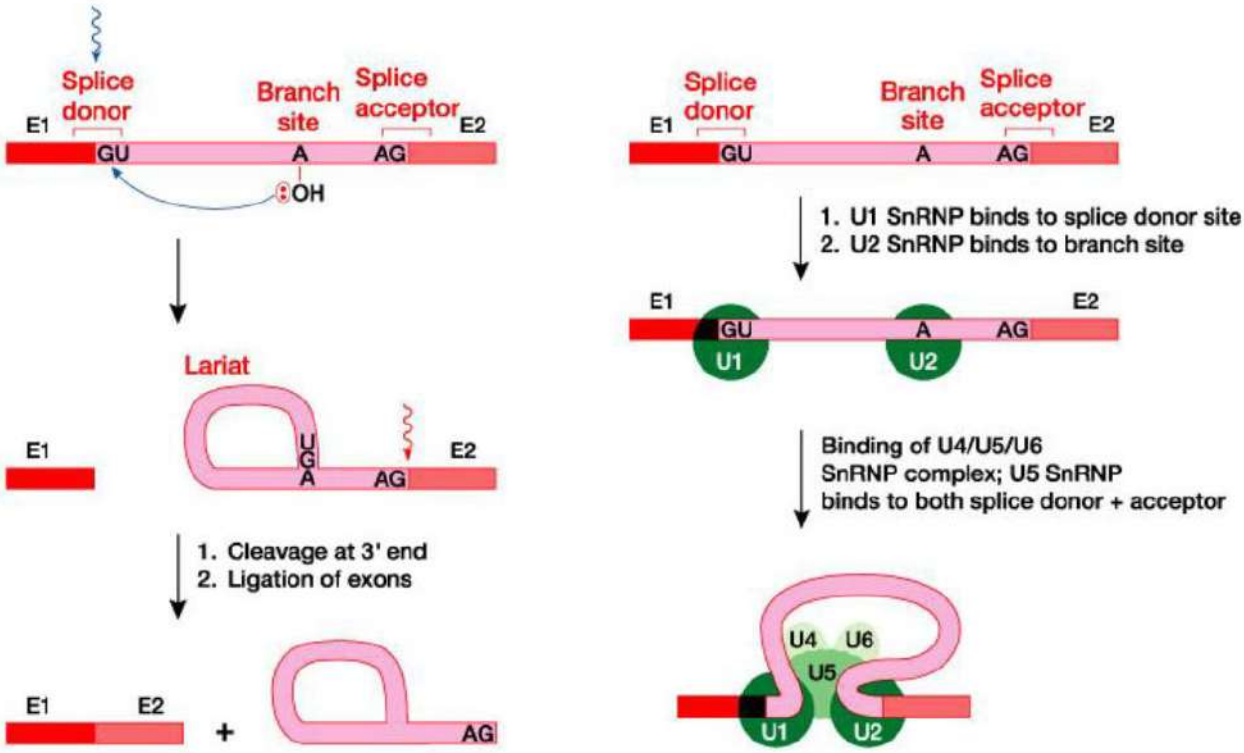
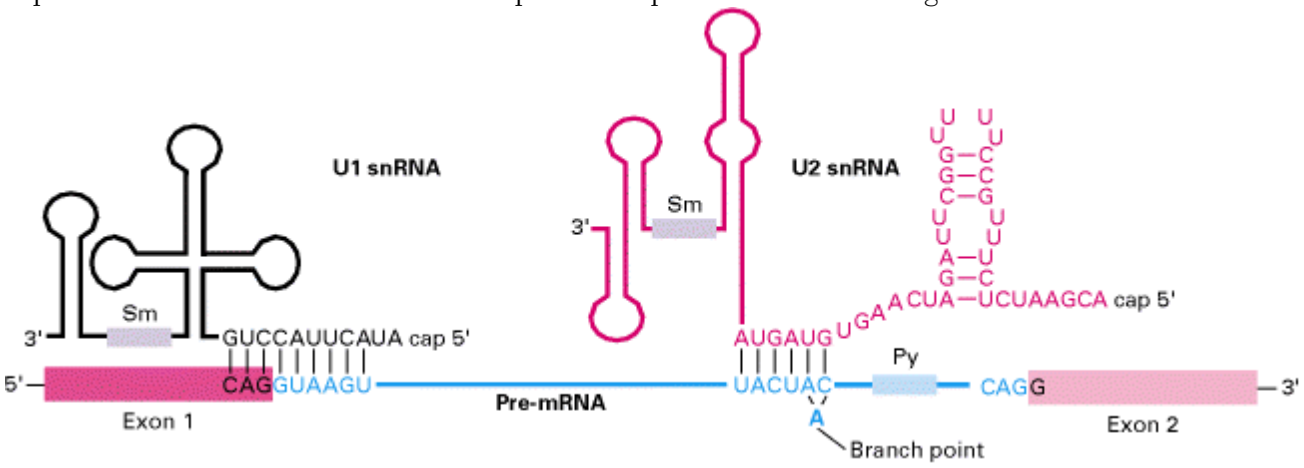


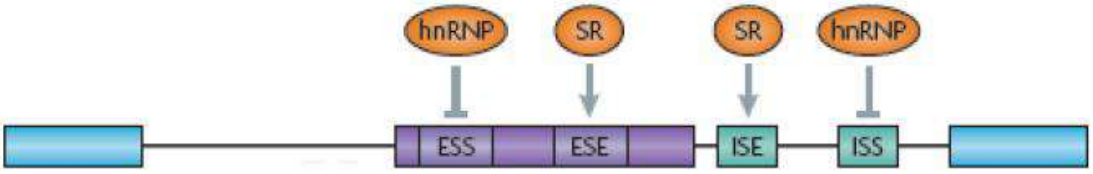
Tabella 14.3 Elenco dei piccoli RNA nucleari (snRNA) dello spliceosoma. Per ciascun snRNA si riporta la lunghezza, l'identificativo della banca dati GenBank e la funzione.

snRNA	Lunghezza (nt)	Genbank ID	Funzione
U1	164	NR_004421	Riconosce il sito di splicing al 5' mediante l'appaiamento di una regione complementare
U2	187	NR_002716	Riconosce il sito di ramificazione mediante l'appaiamento con una regione complementare
U4	144	NR_003925	Forma un duplex con U6
U5	116	X04215	Funzione sconosciuta
U6	106	NR_004394	Forma un duplex con U4; scalza U1 nell'appaiamento con il sito di splicing al 5'

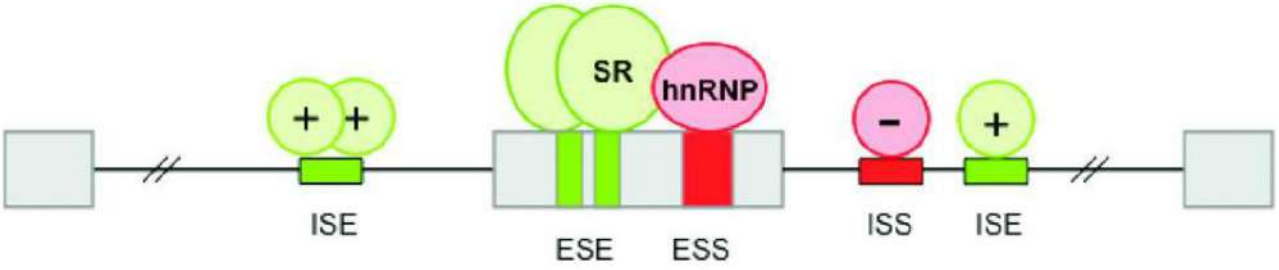
Il posizionamento delle snRNPs è molto preciso in quanto si avvale del legame RNA-RNA.



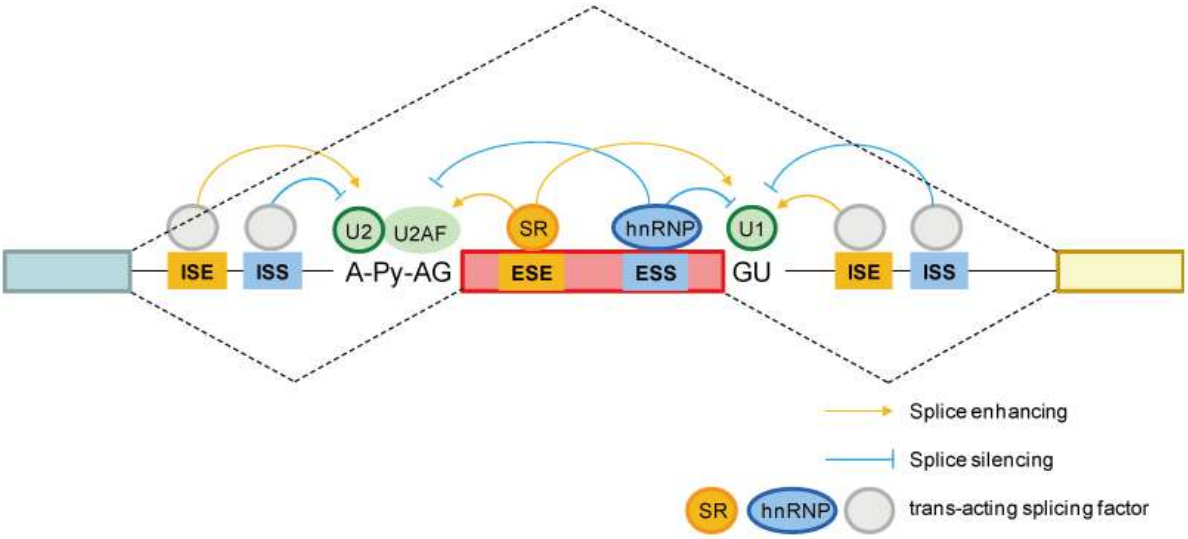
Un altro elemento importante per la regolazione del meccanismo di splicing è la presenza di brevi siti di legame degenerati, su introni ed esoni, per proteine ausiliarie dello splicing. Questi siti sono chiamati enhancers esonici per lo splicing (ESE), enhancers intronici per lo splicing (ISE), silenziatori esonici dello splicing (ESS) e silenziatori intronici dello splicing (ISS). Il riconoscimento dei siti di splicing è mediato da proteine, come le proteine SR o le hnRNPs, che legano specifiche sequenze regolatrici.



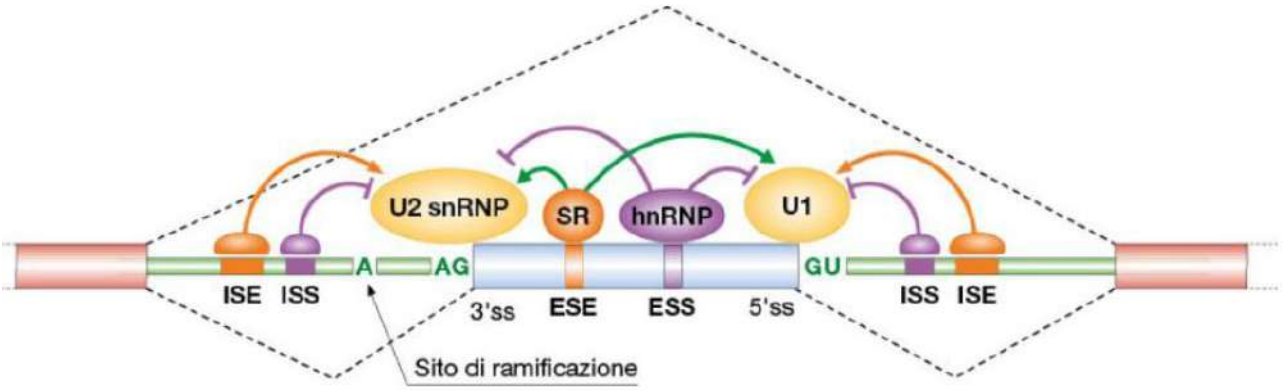
Le hnRNPs sono le heterogenous nuclear ribonucleoproteins, circa 50-100 fattori che si legano alla nascente molecola pre-mRNA e influenzano (negativamente) lo splicing.
Le proteine SR sono proteine altamente conservate che contengono 1 o 2 RNA-recognition motif (RRM) e un dominio C-terminale ricco di residui alternati Serina/aRginina (SR domain).



Le SR e hnRNP sono tessuto-specifiche o stadio-specifiche, e determinano la trascrizione o la inibiscono a seconda delle proteine richieste dalla cellula in cui sono localizzate.



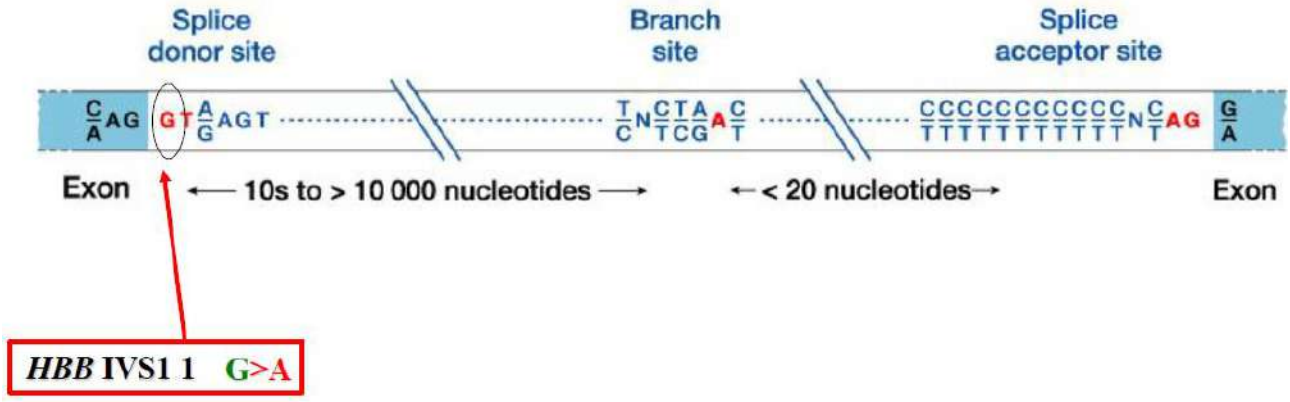
Uno splicing regolato in maniera alternativa da elementi in cis (ESE, ISE, ESS, ISS) ed elementi in trans (proteine SR, hnRNP e simili).



La regolazione dello splicing comporta l'interazione tra proteine in grado di legare specifici motivi presenti nell'mRNA sia negli esoni (ESE, ESS) che negli introni (ISE, ISS). All'azione delle proteine SR che legano le sequenze enhancer negli esoni (ESE) e negli introni (ISE), e che attivano lo splicing (frecce rosse e verdi a punta), si oppongono le proteine hnRNP, che invece riconoscono le sequenze silencer negli esoni (ESS) e negli introni (ISS), indicate dalle freccie viola a punta piatta. Il meccanismo di attivazione e repressione normalmente comporta l'interazione con altre componenti dello spliceosoma.

E' stato osservato che le mutazioni localizzate in elementi in cis e in trans, che danneggiano il codice di splicing o il "macchinario" necessario per lo splicing e la sua regolazione, sono responsabili di varie malattie. E' il caso degli errori di splicing che causano emoglobinopatie.

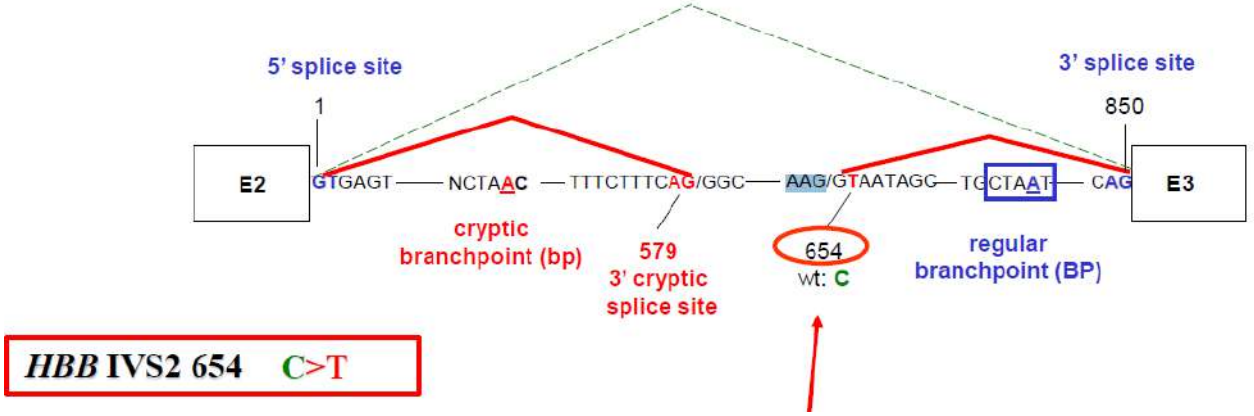
Possono essere mutazioni che colpiscono i confini introne-esone, disattivando i normali siti donatori-accettori di splicing (5' GU – 3' AG), come la G>A ("da G ad A"), alla posizione 1 dell'introne 1 (frequente nelle zone bagnate dal Mar Mediterraneo).



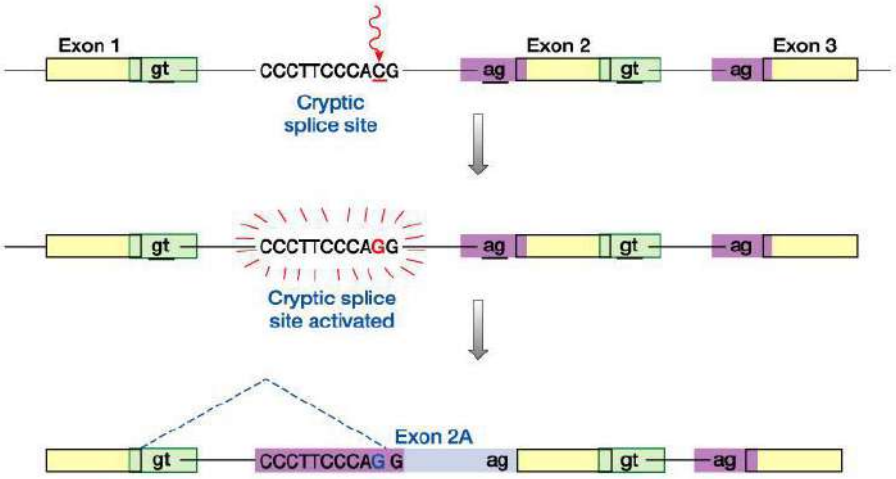
La RNA polimerasi II, non riconoscendo il sito donatore di splicing, ne cerca uno più a valle e, finchè non ne trova uno, continua a trascrivere tutte le basi che legge, producendo un esone aberrante, di lunghezza superiore a quello che si dovrebbe formare. La proteina che si forma risulterà dunque non funzionale.

Altre mutazioni introniche, sempre a carico del gene della β-globina, sono localizzate in posizione 654, 705 e 745 dell'introne 2. Quella alla base numero 654, in particolare, è un C>T, e attiva un sito donatore di splicing criptico, cioè che normalmente non si forma, perché uno previsto c'è, ma è più a monte. Questo sito "nascosto", se si attiva, costringe la RNA polimerasi II, che ha già un sito donatore da cui è partita, alla ricerca del sito accettore corrispondente, a compiere lo splicing che stava portando avanti prima di questo sito criptico appoggiandosi a un sito di ramificazione e uno accettore criptici, che

in condizioni non patologiche avrebbe "saltato", in modo da essere nuovamente libera per ripartire da questo nuovo sito donatore. Si verifica perciò un doppio splicing a carico dell'introne, con la ritenzione, cioè l'inglobazione nell'RNA messaggero maturo di una porzione di introne che andrebbe esclusa con il resto dell'introne. Anche in questo caso il trascritto sarà non funzionante.



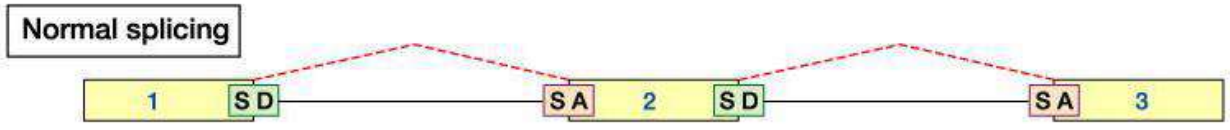
La linea tratteggiata indica lo splicing previsto, quella in rosso mostra lo splicing alternativo che causa alcune emoglobinopatie.



Un terzo tipo di mutazione provoca l'anemia mediterranea β + IVSI-110 G>A, nella quale la mutazione crea un nuovo dinucleotide AG che viene riconosciuto in maniera preferenziale nello "splicing" e così ~80-90% delle molecole di pre-mRNA maturano in modo abnorme e solo 10-20% diventano molecole di mRNA normale. Il sito criptico accettore che si attiva provoca un taglio precoce dell'introne 1, causando un allungamento non previsto dell'esone 2, che viene rinominato, in questi casi patologici, esone 2a.

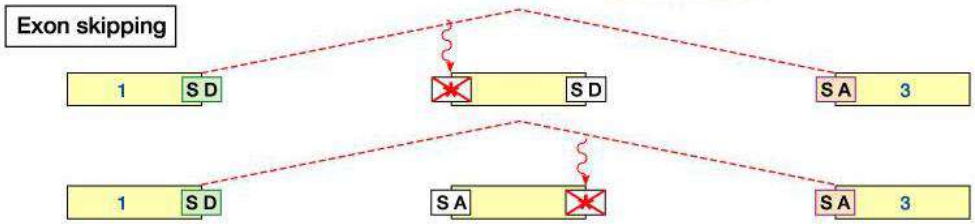
Il nome della mutazione è IVS1 110 (introne 1 posizione 110).

A partire da una situazione di splicing normale,

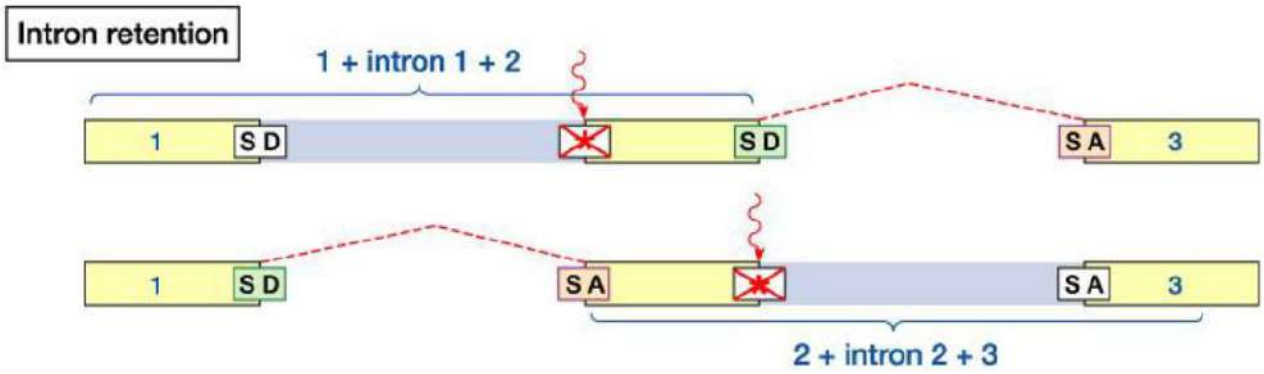


gli effetti delle varianti di splicing sono quindi fondamentalmente di quattro tipi:

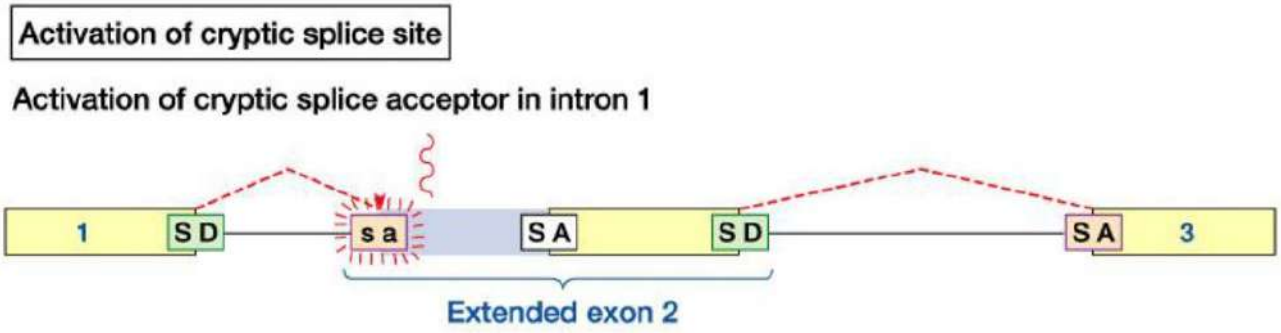
- ✓ Esclusione dell'esone dall'mRNA (exon skipping);



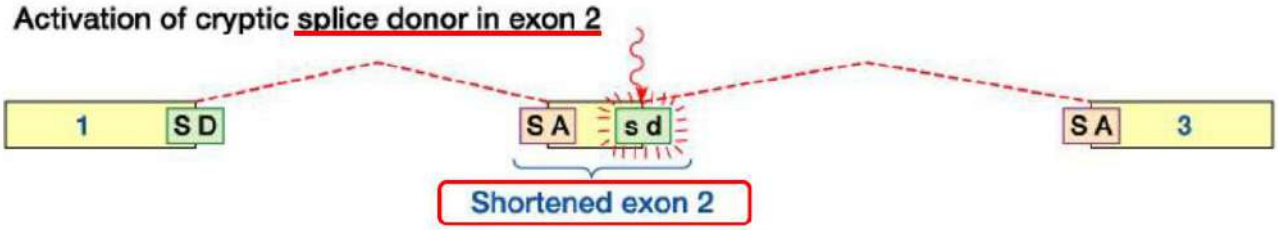
✓ Ritenzione dell'introne (intron retention): l'introne non viene eliminato nel corso dello splicing;



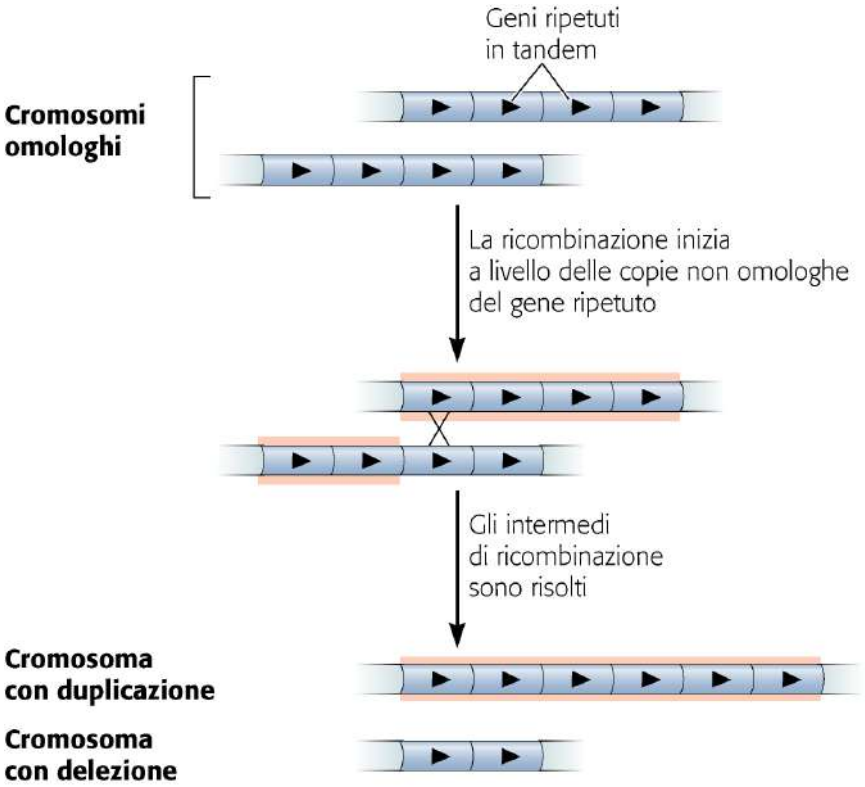
✓ Inclusione di una parte dell'introne nel trascritto maturo, con conseguente allungamento dell'esone seguente;



✓ Esclusione di una parte di esone dall'mRNA, che risulta nell'accorciamento dell'esone colpito.

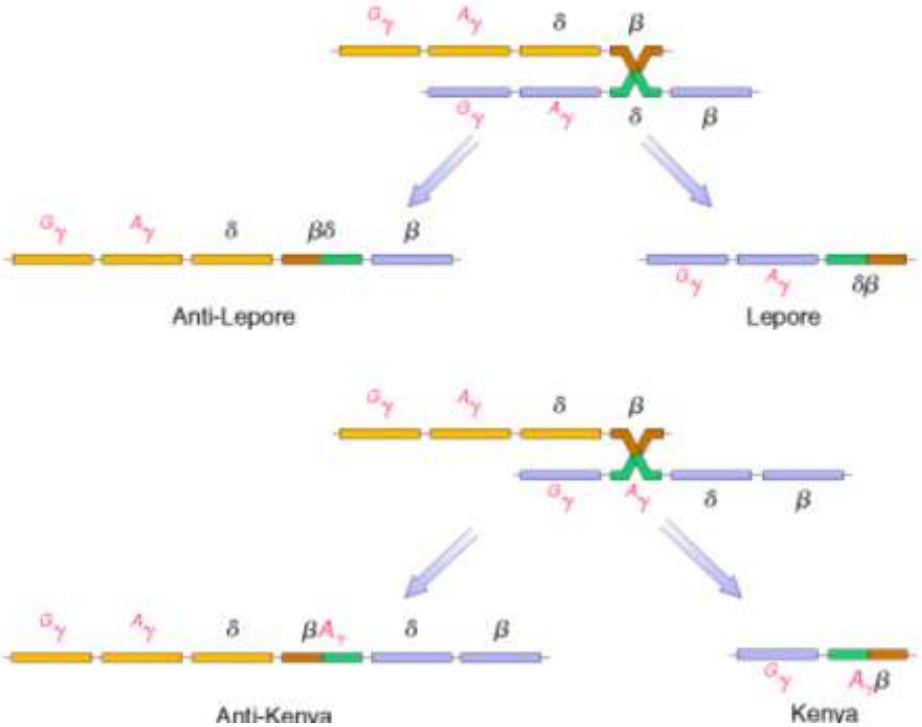


Le mutazioni missenso, ovvero le sostituzioni di un amminoacido normalmente presente nella proteina wild-type con un altro, possono provocare emoglobinopatie. In particolare, una mutazione nel codone numero 6, GAG, codificante per il glutammato, in un codone GAU, codificante per la valina, impedisce il normale sviluppo del polipeptide β normale, facente parte del complesso dell'emoglobina α, provocando la formazione di polipeptidi falciformi, dove l'emoglobina viene classificata come "S".
Possono anche verificarsi mutazioni nonsenso o stop-gain, che determinano l'inserimento di un codone di stop al posto di un amminoacido, provocando la prematura fine della traduzione e la sintesi di una proteina incompleta perché accorciata, non funzionale, che deve essere degradata dalla cellula, in quanto dannosa per l'organismo. A livello del codone 15 una tale mutazione è diffusa in India, a livello del 17 in Cina, a livello del 39 nella zona mediterranea.
Il gene dell'α globina normale presenta una sequenza terminatrice (UAA) a livello del codone 142, che segna la fine della traduzione. Se questa tripletta muta in GAA (codone della glutammina), ovvero si osserva una mutazione "read-through" ("continua a leggere"), la proteina viene terminata più a valle di una trentina di basi. L'emoglobina sintetizzata, patologica, prende il nome di CS Hb, dove CS sta per Constant Spring. E' associata alla malattia da emoglobina H.
Un ultimo tipo di variante per i geni delle globine è dovuto al crossing-over ineguale.



Il crossing-over diseguale (o ricombinazione omologa non allelica – non homologous allelic recombination – NHAR) si realizza tra sequenze non alleliche di cromatidi non fratelli o cromatidi fratelli. Questo fenomeno porta alla genesi di inserzioni/delezioni (delezione su un cromatidio, inserzione sull'altro), che possono essere patogene perchè causano la perdita o l'acquisizione di informazione genetica, il cui effetto finale può essere quello di alterare la struttura o gli equilibri stechiometrici di alcune proteine.

Alla base del fenomeno del crossing-over vi è la presenza di sequenze ripetute. Lo scambio tende cioè a verificarsi in corrispondenza delle regioni cromosomiche dove vi è un alto grado di ripetitività delle sequenze (tali regioni sono rappresentate in gran quantità nel genoma umano). Il NHAR in particolare è causato dall'appaiamento sfasato fra sequenze ripetute su cromatidi fratelli o non fratelli.

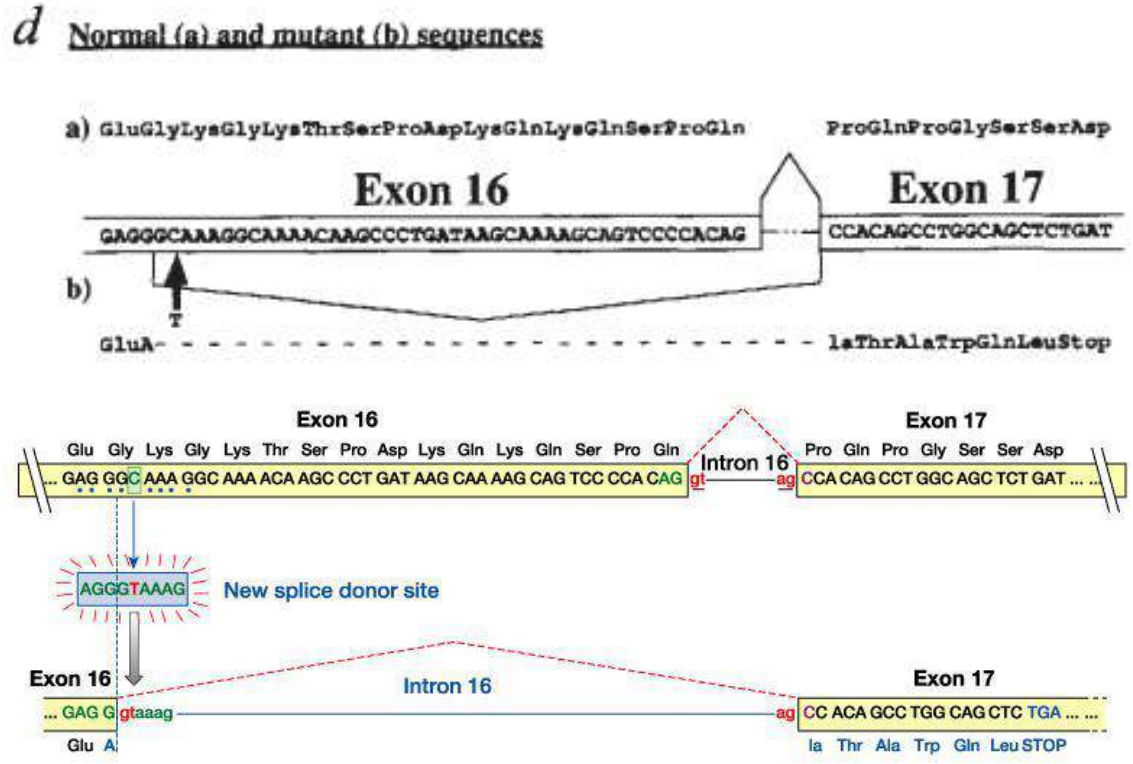


Due forme di emoglobine che si originano in seguito a crossing-over ineguale. L'emoglobina Anti-Lepore è nota anche con il nome di "Hb Miyada".

La frequenza delle varie mutazioni che colpiscono i geni dell'emoglobina sono diverse a seconda della popolazione.

Alcune delle mutazioni più deleterie per la proteina nascente sono però quelle frameshift. Una mutazione frameshift è dovuta a inserzione o delezione di una o poche coppie di basi. Come risultato si ha uno scorrimento del modulo o cornice di lettura (frame) dal sito mutato in poi. Cioè il ribosoma, il sito della traduzione, da un certo punto in avanti, sintetizza, in moltissimi frangenti, amminoacidi completamente diversi da quelli w-t, sintetizzando una proteina abnorme, che viene subito degradata dalle proteasi cellulari.
A proposito della ORF (Open Reading Frame) è opportuno ricordare che a livello del DNA gli introni possono interrompere una sequenza codificante in una di tre possibili posizioni: tra due codoni, fra il primo e il secondo nucleotide di un codone o fra il secondo e il terzo nucleotide di un codone.
Un esone interno può essere affiancato da introni che si inseriscono nello stesso tipo di posizione; in un esone come questo (chiamato esone simmetrico), il numero dei nucleotidi è sempre esattamente divisibile per tre. Un esone interno asimmetrico ha un numero totale di nucleotidi non esattamente divisibile per tre e perciò è affiancato da introni che si inseriscono in posizioni differenti.
A volte, però, le mutazioni patogene non sono evidenti: è il caso di mutazioni sinonime, dove la tripletta codifica sempre per lo stesso amminoacido, ma la base mutata influirà sul legame, ad esempio, con i fattori di trascrizione che si legano usualmente a quel sito, che non riconosceranno più il loro "attracco", compromettendo, anche, la funzionalità della proteina stessa.

In questa situazione, la mutazione A->C causa la sostituzione di un codone di arginina (CGA) con un altro codone sempre di arginina (CGC), che è comunque patogeno perché modifica la sequenza enhancer di splicing evidenziata, che si trova all'inizio dell'esone (evidenziato in azzurro).
Un lavoro del 1995 di Richard e Beckmann pubblicato su "Nature genetics" ha chiarito quanto in realtà le mutazioni silenti non fossero così neutre ai fini della trascrizione, come si pensava.
Questa ricerca dimostrò il coinvolgimento della calpaina specifica dei muscoli nell'eziologia della distrofia muscolare dei cingoli, legata al cromosoma 15, compresa l'inaspettata osservazione di almeno cinque mutazioni distinte a carico della calpaina in una piccola comunità isolata di consanguinei dall'isola di La Réunion. Ulteriori mutazioni segregavano in questa popolazione, mentre i pazienti di una famiglia (R15) non portavano nessuna delle mutazioni identificate. Si procedette quindi a uno screening sistematico del gene della calpaina (CANP3) di un paziente di questa famiglia geneticamente omozigote, servendosi di analisi in eteroduplex e screenando ciascun esone del gene. L'unica variazione nella sequenza genomica che venne identificata fu una mutazione apparentemente silente alla terza base nel codone 624 della glicina (GGC->GGT), in posizione 1872 del cDNA. Questo codone sinonimo si sarebbe potuto considerare un polimorfismo neutro. Fu successivamente determinato, però, che questa mutazione aveva carattere patologico in seguito alla visualizzazione di un prodotto di trascrizione di taglia anormale. Il sequenziamento del corrispondente frammento amplificato con la PCR rivelò un prodotto sottoposto ad un processo di splicing aberrante, che aveva provocato la delezione delle ultime 44 basi dell'esone 16.



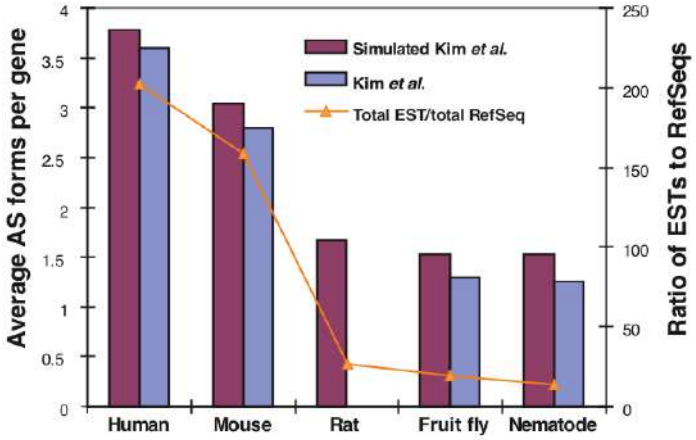
Ma qual è il numero di geni nell'uomo?

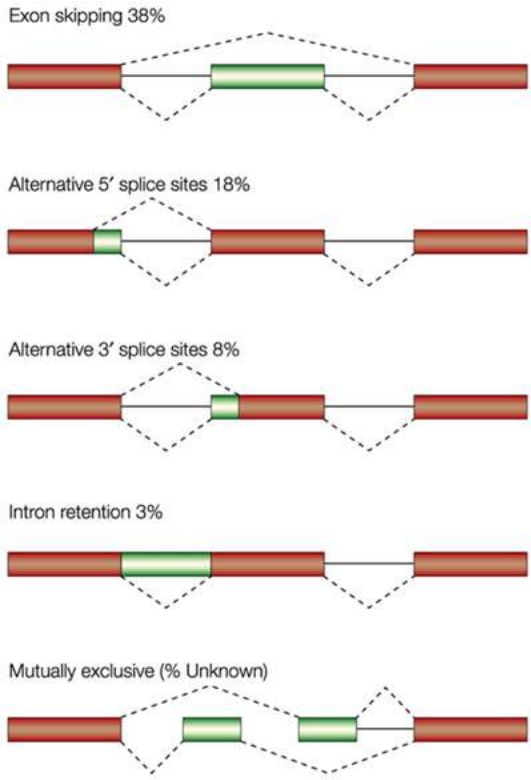
Le stime più recenti e più attendibili indicano un numero che si aggira intorno ai 20 000; nel 2004 si pensava fossero tra i 20 000 e i 25 000; ancora prima, nel 2001, gli stessi scienziati avevano ipotizzato dai 30 000 ai 40 000.

Il nostro genoma deriva un trascrittoma composto da un numero di trascritti compreso tra 50 000 e 250 000, mentre il proteoma umano ha un numero non ancora ben determinato di proteine: non meno di 200 000, non più di un milione.

Lo splicing alternativo

In seguito ai lavori originali del 1977 dove veniva descritto il meccanismo dello splicing del pre-mRNA, si capì rapidamente che lo splicing contemporaneo di diverse combinazioni di siti di splicing, il cosiddetto splicing alternativo, permetteva ai geni di generare più di una isoforma di mRNA per gene (nel caso dell'uomo una media di 3,8 isoforme per gene, mentre negli organismi meno complessi il numero si riduce fino a 1,5). La reale dimensione dello splicing alternativo iniziò ad essere compresa quando furono intrapresi i progetti di sequenziamento del genoma su larga scala e del trascrittoma, portando alla constatazione che lo splicing alternativo è la regola piuttosto che l'eccezione.



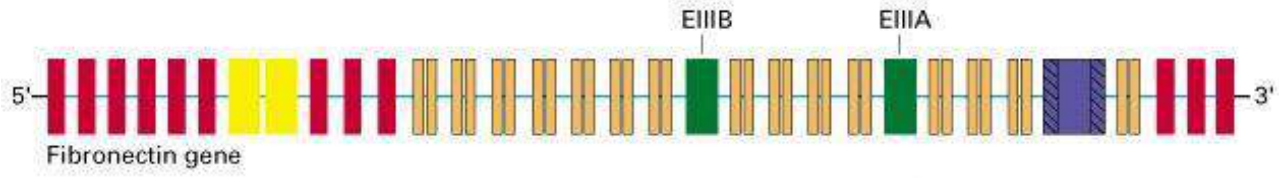


La grande maggioranza dei geni multiesonici nell'uomo subisce lo splicing alternativo, che aumenta enormemente la diversità funzionale delle specie proteiche.

Anche in altri organismi pluricellulari questo avviene in gran parte dei casi, ma nell'uomo la percentuale (molto di più di quanto si aspettassero gli scienziati) raggiunge quasi il 90% dei geni codificanti per proteine.

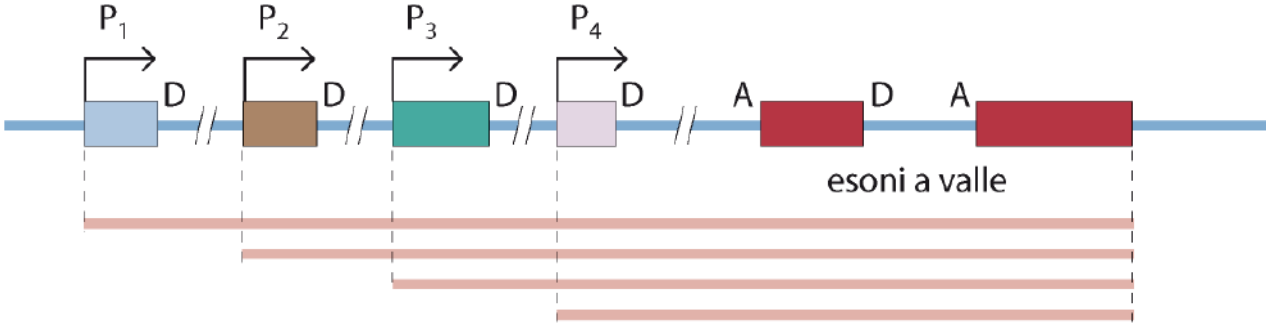
Un caso di splicing alternativo molto noto è quello che coinvolge il gene della fibronectina, una glicoproteina dimerica prodotta da molte cellule e tessuti, presente nel tessuto connettivo lasso e in quello denso.

Il suo gene presente due domini esonici (EIIIA e EIIIB), che vengono espressi, cioè inclusi nel trascritto maturo, soltanto, ad esempio, nei fibroblasti, dove la proteina deve essere aderente alla superficie cellulare. Negli epatociti, invece, la cellula non ha questa necessità e deve invece possedere una fibronectina solubile, quindi l'RNA polimerasi skippa i due domini, che non vengono, perciò, trascritti.



I promotori alternativi

Un lavoro del 1996 aveva evidenziato come l'uso di promotori multipli (contemporaneamente) e di siti di inizio di trascrizione fosse apparentemente, fino a quel momento, frequente. Nello stesso tempo, però, si osservò anche una considerevole variazione e complessità negli schemi di utilizzo di promotori alternativi. Era la prima volta che si descriveva l'uso di promotori alternativi come meccanismo versatile responsabile di diversità e flessibilità nella regolazione dell'espressione genica.



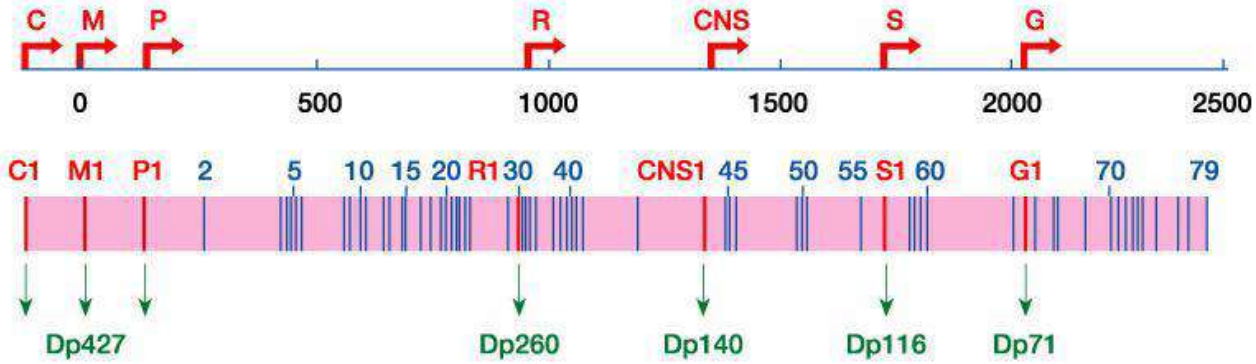
In questo gene contiamo 4 possibili promotori alternativi, con altrettanti possibili primi esoni e trascritti primari alternativi. La dimensione media del primo introne è di 14 200 bp contro una di 4 800 bp dei seguenti esoni. I promotori alternativi sono responsabili di diverse isoforme, legate a diversi fattori:

- ✓ Specificità di tessuto (es. il gene della distrofina). Anche se molti geni che mostrano un'espressione tessuto-specifica vengono spesso trascritti a livelli molto bassi in tutte le cellule (trascrizione illegittima o ectopica);

- ✓ Specificità per uno stadio di sviluppo (es. i geni per i fattori di crescita);
- ✓ Localizzazione intracellulare differenziata (es. un'isoforma solubile ed una legata alla membrana, cfr. esempio precedente della fibronectina).

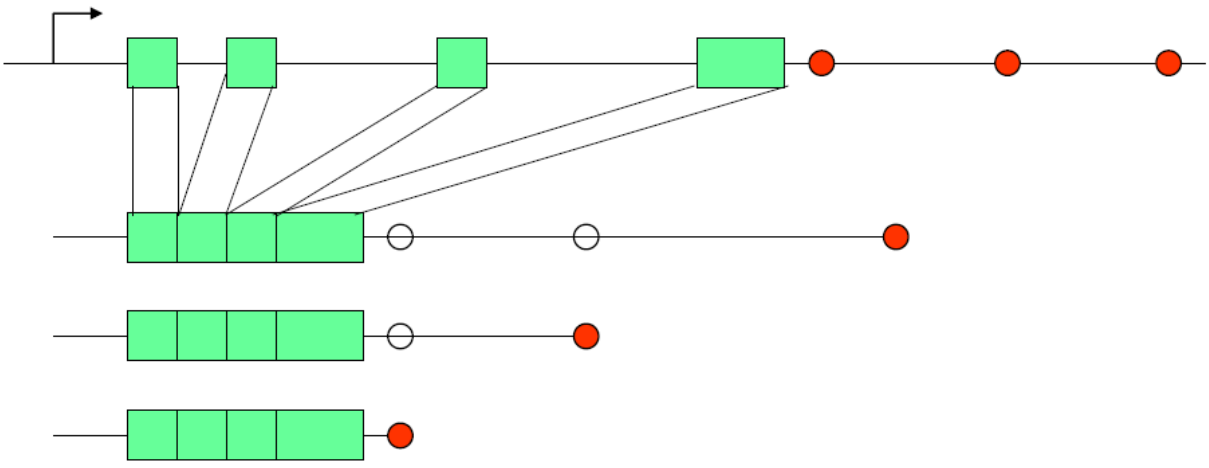
Oggi possiamo affermare che i geni con promotori alternativi sono comuni ai genomi dei Mammiferi. Prove raccolte recentemente mostrano che il 30-50% dei geni umani e circa la metà di quelli del topo possiedono diversi promotori alternativi, che possono estendersi anche per migliaia di kilobasi. L'uso aberrante di un promotore rispetto ad un altro è stato trovato essere associato a diverse malattie, compreso il cancro.

Uno degli esempi più noti è il gene della distrofina, appunto, che presenta ben sette promotori alternativi tessuto-specifici differenti.



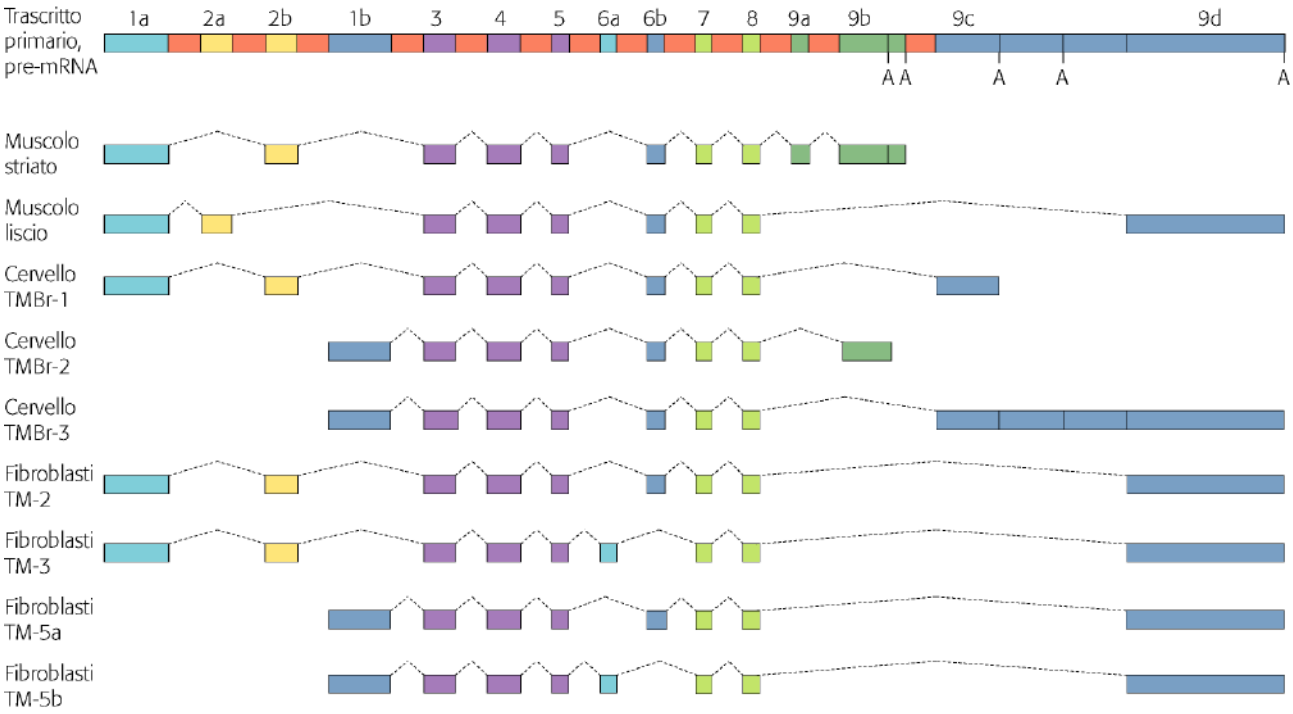
La poliadenilazione alternativa

Infine, un altro meccanismo piuttosto comune con il quale possono essere generati trascritti diversi, è la poliadenilazione alternativa. Molti geni, infatti, presentano nelle regioni 3' due o più segnali di poliadenilazione che possono essere utilizzati durante la maturazione degli RNA. Anche in questo caso si possono formare isoforme tessuto-specifiche.



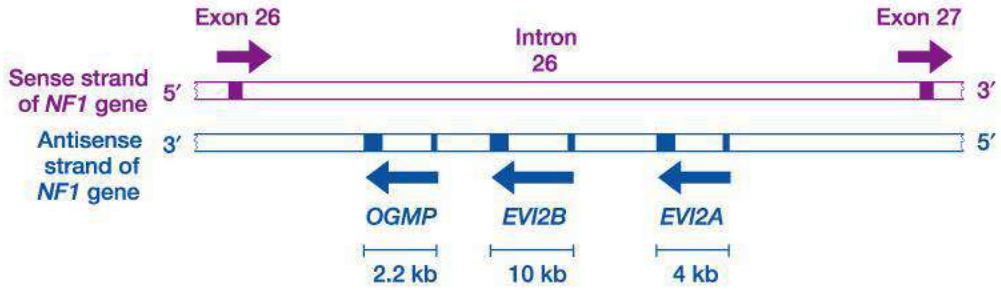
In effetti, il segnale di poliadenilazione (AAUAAA) talvolta è "saltato" dal complesso che taglia e aggiunge le code di poli(A) ai messaggeri. Questo genera variabilità nelle sequenze 3' UTR degli RNA messaggeri.

In molti casi, però, durante la maturazione di uno specifico RNA, può verificarsi una combinazione di questi meccanismi, che permette la formazione di numerose isoforme proteiche a partire da un singolo gene. Ciò avviene, ad esempio, per il gene dell'α-tropomiosina, per il quale si combinano splicing e poliadenilazione alternativi.



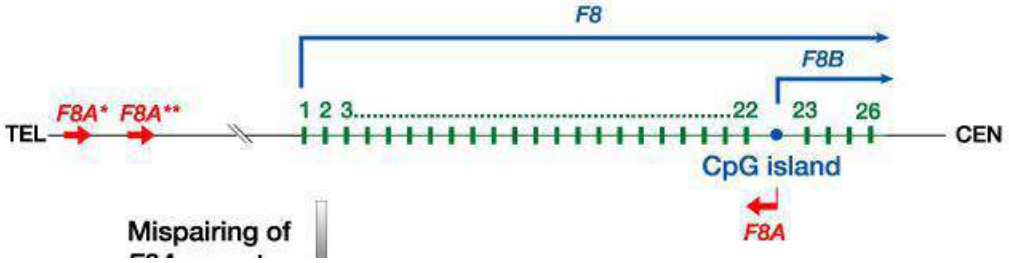
I geni nested

Tra i 20 000 geni che codificano per proteine nel nostro organismo, esistono alcuni chiamati geni nested o "annidati". Costituiscono il 10% del totale dei geni e vengono così denominati perché si sovrappongono ad un altro gene. Il 90% delle sovrapposizioni coinvolge geni trascritti da filamenti opposti. I primi geni nested erano stati identificati nei batteri e nei virus.

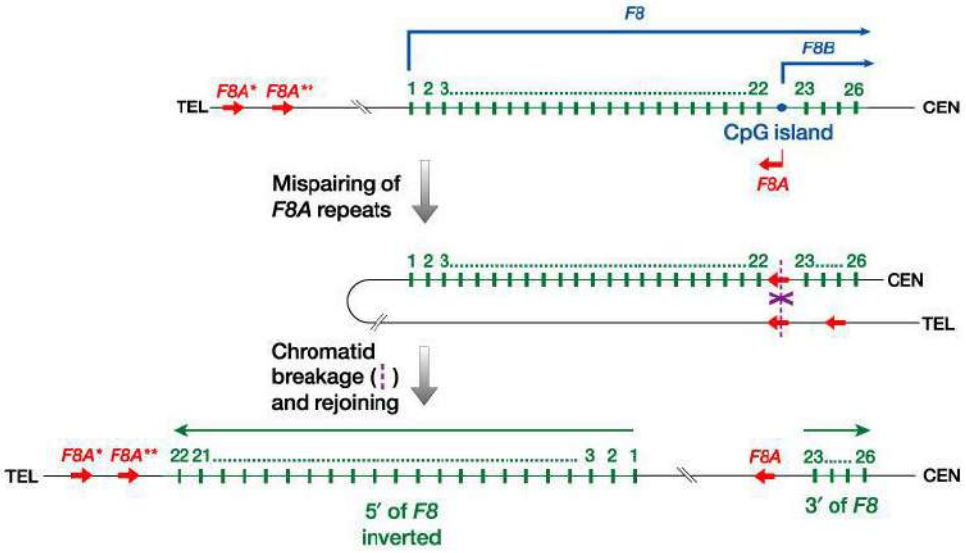


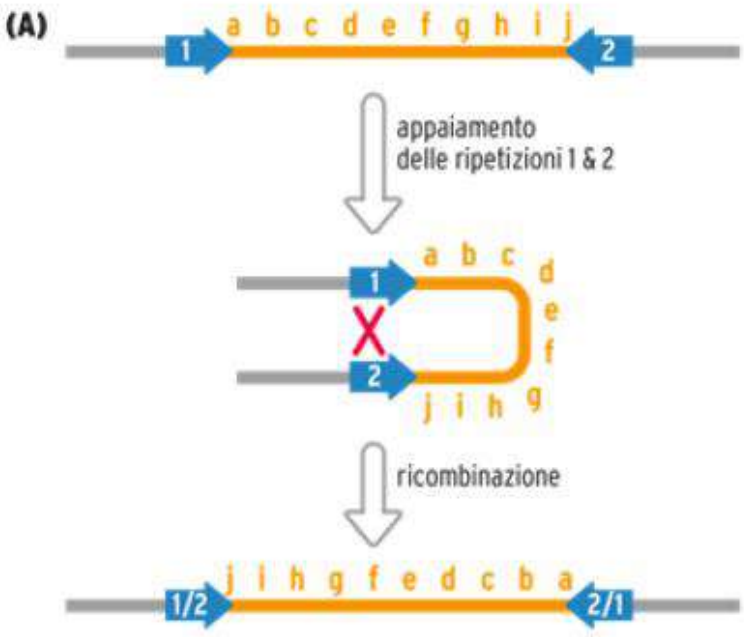
Un esempio di geni codificanti all'interno di altri geni si osserva per il gene NF1 (neurofibrominina1). Sul filamento senso la RNA polimerasi II trascrive l'RNA messaggero di NF1, mentre dal filamento antisenso trascrive ben tre geni: OGMP, per la glicoproteina mielinica oligodendrocitica, e EVI2B e EVI2A, omologhi umani di geni murini coinvolti nella genesi della leucemia.

Un altro caso notevole di trascrizione bidirezionale è a carico del gene del fattore VIII della coagulazione, che codifica sul cromosoma X. Questo gene è intronless, dunque è composto da soli esoni. Tra l'esone 22 e il 23 si trova un' "isola CG", alla quale si lega un promotore, che esplica la proprià attività bidirezionalmente, sullo stesso filamento.

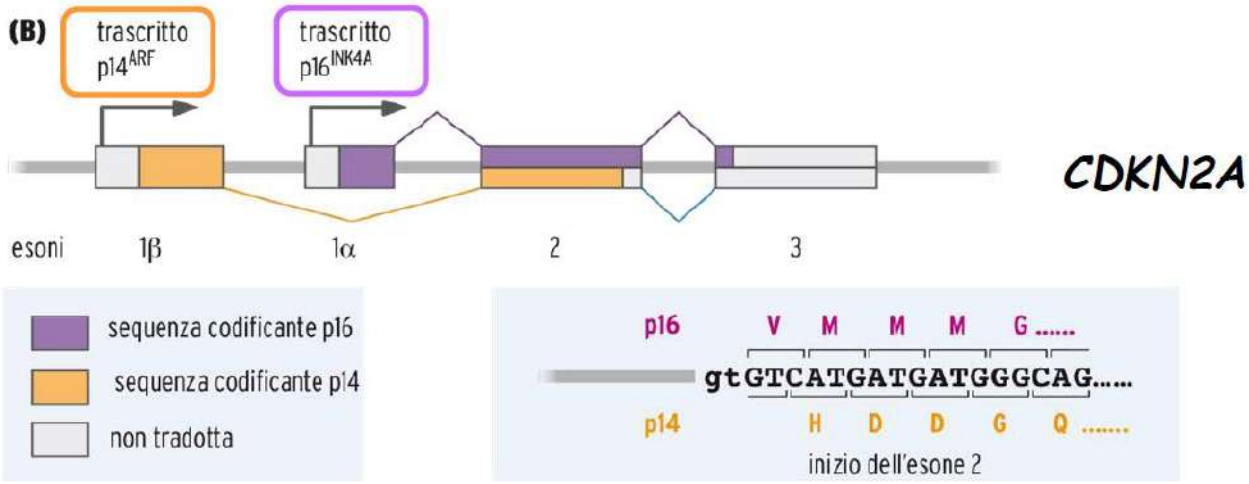


A 400-500 kb a monte della regione esonica, verso il telomero del cromosoma, sono presenti F8A* e F8A**, due sequenze strettamente correlate a F8A. La grande omologia di sequenza tra le varie F8A determina l'appaiamento scorretto di queste sequenze ripetute, provocando l'associazione di F8A con F8A**, che, avendo orientamento opposto, sono responsabili del ripiegamento del DNA su se stesso. seguita da una sorta di crossing-over intracromatidico. La rottura della doppia elica in corrispondenza delle sequenze ripetute (con conseguente risoluzione del chiasma) e la successiva riparazione determinano l'inversione dell'orientamento del frammento di DNA compreso tra i due punti di rottura, come dimostrato da Lakich et al. nel 1993. Si parla, infatti, di inversione intragenica, un evento relativamente raro che avviene a causa della presenza di sequenze duplicate e invertite (F8A e F8A** hanno direzioni opposte), localizzate all'interno di un gene o in prossimità di esso. L'inversione del fattore VIII provoca l'insorgenza dell'emofilia del tipo A.

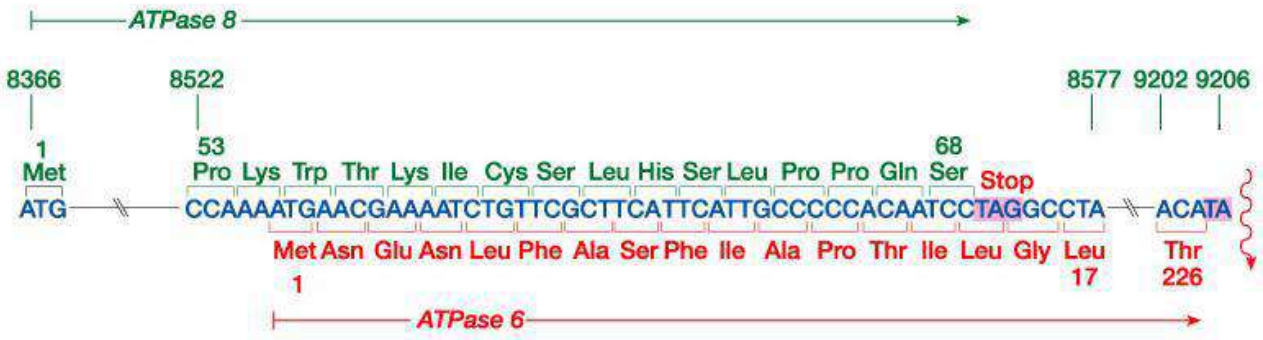




Il 99% dei casi di inversione del fattore VIII si originano alla meiosi del nonno materno. Nella spermatogenesi dei mammiferi, infatti, i cromosomi sessuali vengono segregati nel corpo "XY" e quindi sono compattati in stati eterocromatinici, che piegano l'estremità telomerica verso il centromero, accorciando la lunghezza del cromosoma.
La presenza di geni nested può accompagnarsi a fenomeni come lo splicing alternativo. E' il caso del gene CDKN2A (inibitore della chinasi ciclina-dipendente 2a), che codifica per due proteine importanti per il corretto svolgimento del ciclo cellulare, attraverso l'utilizzo di regioni codificanti condivise e cornici di lettura alternative: p16(INK4), che è un inibitore della chinasi ciclina-dipendente, e p14(ARF), che lega la proteina in grado di stabilizzare p53, MDM2. Di questo gene sono coinvolti, in diversa misura, i primi tre esoni. I due trascritti saranno, perciò, profondamente diversi.



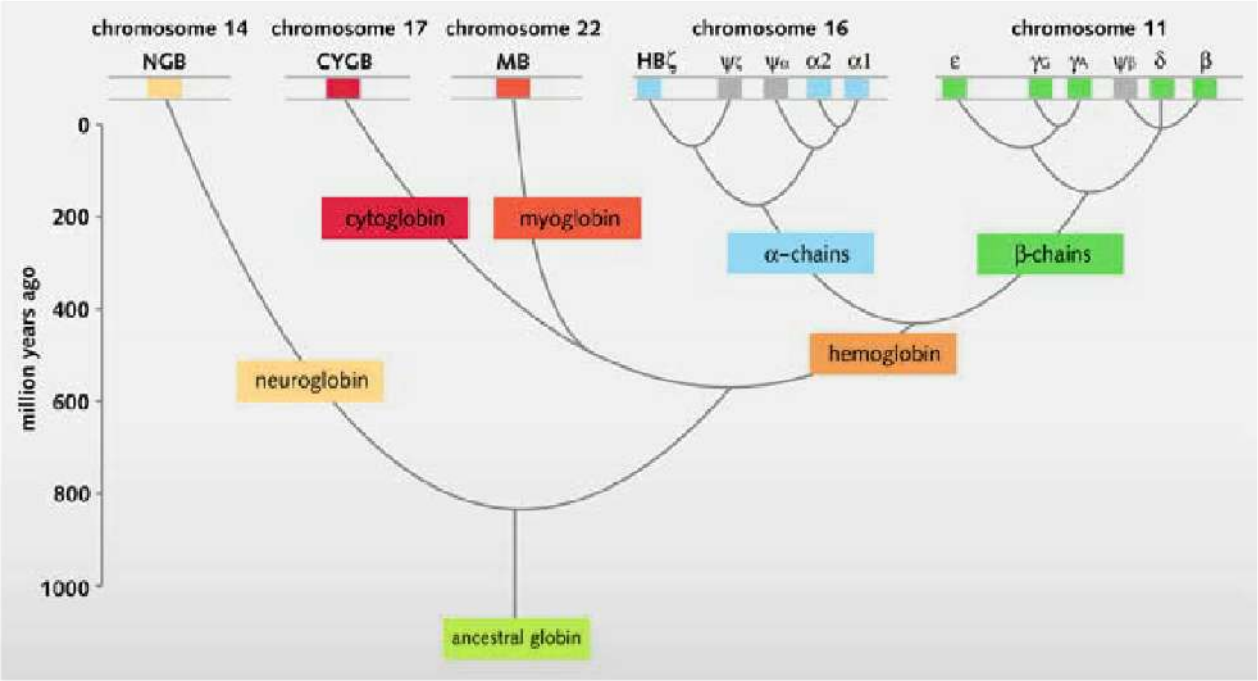
Nei DNA circolari di alcuni microrganismi, la stessa porzione genomica codifica, ad esempio per l'ATPasi 6 e per la 8.

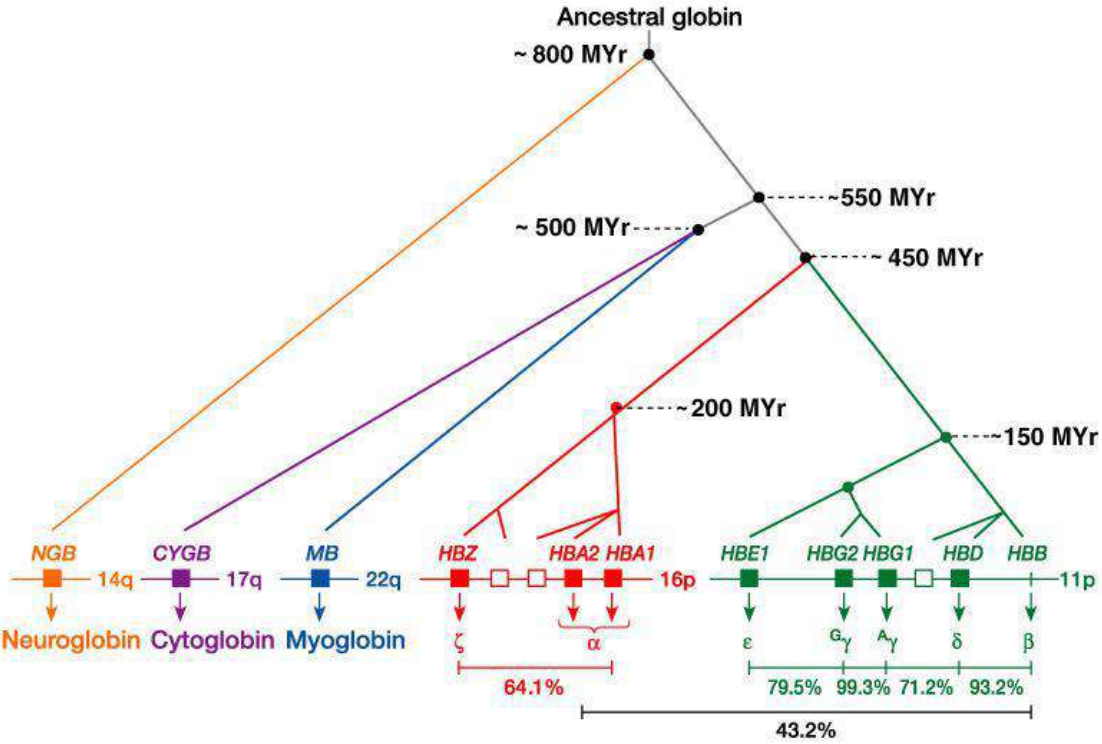


Evoluzione delle famiglie multigeniche per duplicazione genica

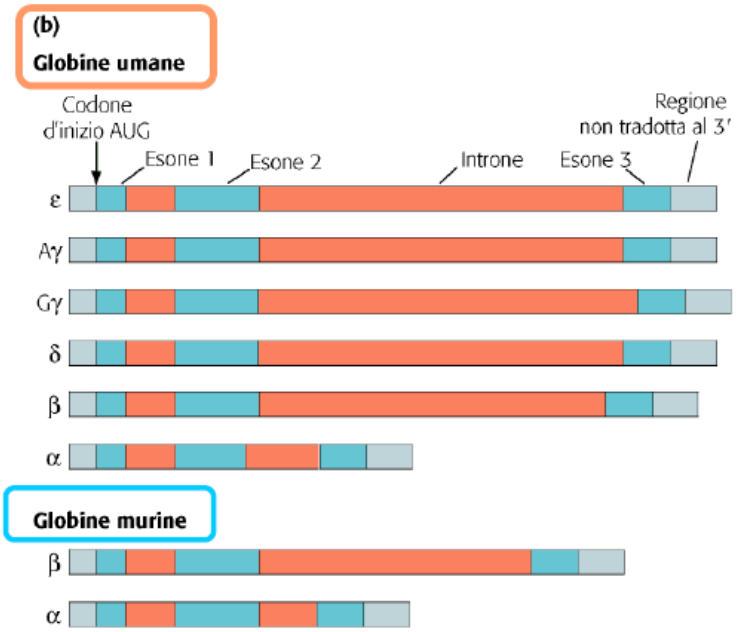
Negli eucarioti frequentemente si riscontrano copie multiple di geni, tutte con sequenza identica o simile. Un gruppo di tali geni è detto famiglia multigenica, che corrisponde a un insieme di geni correlati che si sono evoluti da un certo gene ancestrale mediante un processo di duplicazione genica. I membri di una famiglia genica possono trovarsi raggruppati oppure dispersi su cromosomi diversi. Il sequenziamento dei genomi ha rivelato che le famiglie multigeniche sono componenti comuni di tutti gli organismi. Numerosi articoli su questo argomento vennero pubblicati sulla rivista "Natural Genetics" nel 2002, visto all'interno del phylum dei Cordati, del subphylum dei Vertebrati, ad esempio, confrontando i geni dell'uomo con quelli di altre specie caratterizzate da una struttura scheletrica ossea e/o cartilaginea.

Un esempio di famiglia multigenica è rappresentato dai geni della globina, che codificano per le catene polipeptidiche costituenti la molecola dell'emoglobina. Nell'uomo, la famiglia multigenica della globina è composta da 7 geni a-simili sul cromosoma 16 e da 6 geni b-simili sul cromosoma 11. I geni per la globina si trovano anche in altri animali, e geni globina-simili vengono trovati persino nelle piante, per cui dovrebbe trattarsi di una famiglia genica molto antica.
Per rappresentare le famiglie multigeniche in modo più efficace dal punto di vista grafico si ricorre spesso a diagrammi "cluster": ramificazioni evolutive che partono da un progenitore ancestrale e che spiegano la diversificazione genica che si è verificata nel corso del tempo. Questo è il cluster della globina.



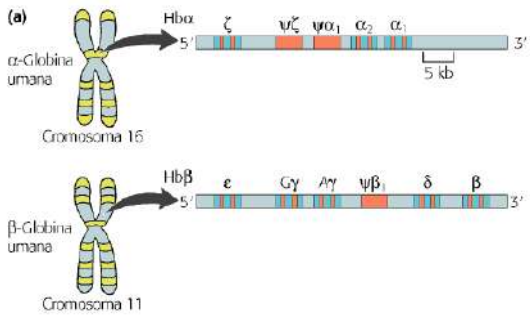


Circa 500 milioni di fa deve essere avvenuta una duplicazione genica, che ha portato alla specificazione di una copia di globina in neuroglobina. Sul meccanismo con cui si possono verificare le duplicazioni torneremo più avanti. Nel corso dell'evoluzione la globina "parentale" è andata incontro più volte a duplicazioni e a conseguenti divegenze, che hanno reso ciascun "figlio" genico unico, seppure con sequenza simile a geni "cugini" già differenziati o ancora da sviluppare. La divergenza è stata resa possibile dall'accumulo di mutazioni, che, in maniera puntiforme ma costante, hanno lentamente alterato la sequenza originaria, senza stravolgerne il pattern di partenza. Dal momento che tutti i geni globinici hanno similarità di sequenza e di struttura, sembra che un gene globinico ancestrale (forse correlato al gene attuale per la mioglobina) si duplicò e produsse per divergenza un gene a-simile ancestrale e un gene b-simile ancestrale. Questi due geni andarono incontro a ripetute duplicazioni, dando luogo ai vari geni a-simili e b-simili presenti nei vertebrati odierni. I geni per mioglobina e citoglobina vengono definiti, tra di loro, paraloghi, perché appartengono a un gruppo di sequenze omologhe (derivanti dalla globina, in questo caso) all'interno di una stessa specie. Con l'evoluzione e la comparsa di nuove specie, la speciazione, i geni, in ogni essere vivente, si sono distribuiti indipendentemente in un organismo rispetto all'altro. Il gene della mioglobina nell'uomo e il medesimo gene nel topo vengono definiti, reciprocamente, ortologhi. Lo stesso discorso vale per quello della citoglobina, per esempio.



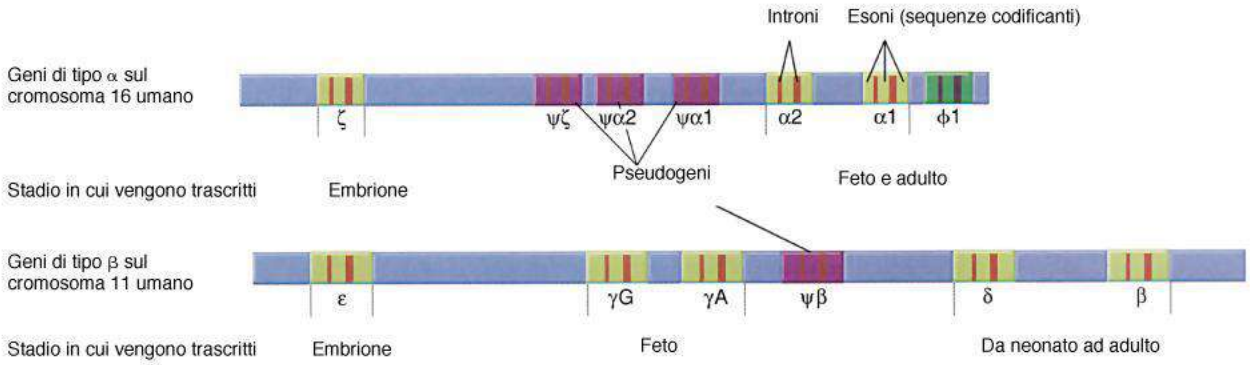
Nelle specie animali quasi tutti i geni globinici funzionali possiedono la stessa struttura generale, che consiste in 3 esoni separati da due introni. Qui, in figura, vengono confrontati i geni globinici umani con quelli murini.

Nell'uomo, la famiglia multigenica della globina è composta da 7 geni a-simili sul cromosoma 16 e da 6 geni b-simili sul cromosoma 11.

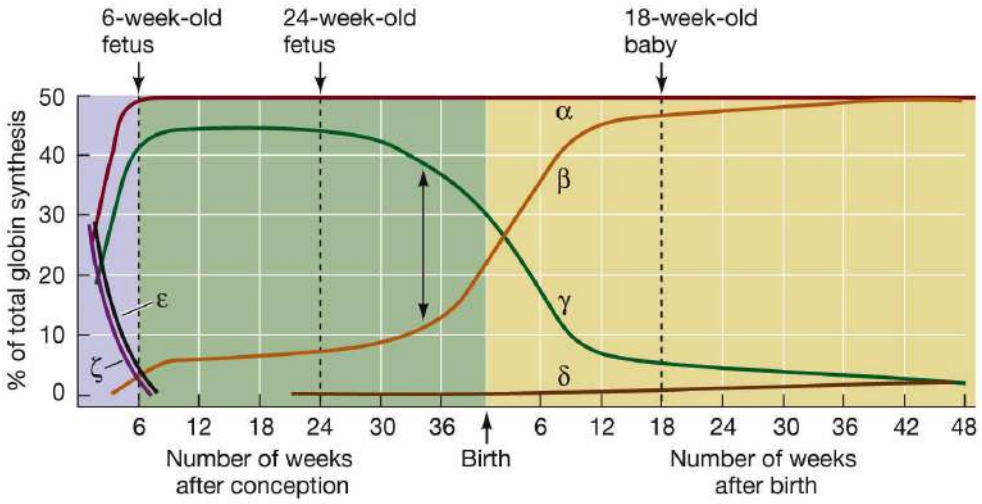


Duplicazioni geniche ripetute, come quelle che hanno dato luogo alla famiglia genica per la globina, sembrano un avvenimento evolutivo frequente e, in effetti, il numero di copie dei geni globinici varia in alcune popolazioni umane. Ad esempio, la maggior parte degli uomini possiede due geni per l'a-globina sul cromosoma 16; tuttavia alcuni individui presentano sul cromosoma 16 un unico gene per l'a-globina, mentre altri individui presentano, su uno dei loro cromosomi, tre o anche quattro copie del gene per l'a-globina. I geni per la β-globina codificano invece sul cromosoma 11.

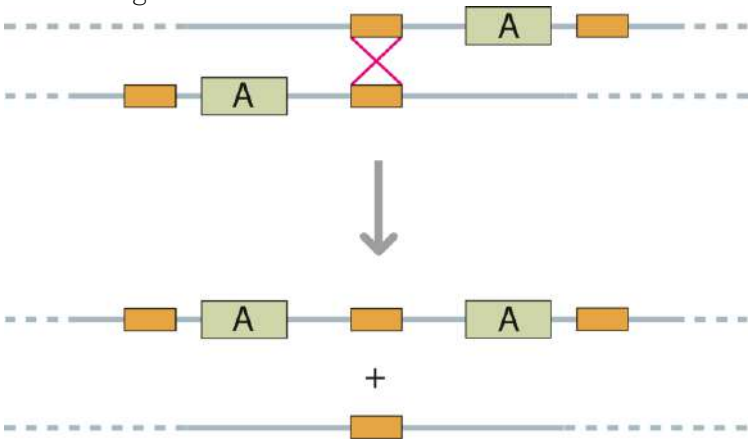
Nel corso dello sviluppo umano, in momenti diversi e in tessuti diversi vengono espressi geni differenti della famiglia genica della globina.



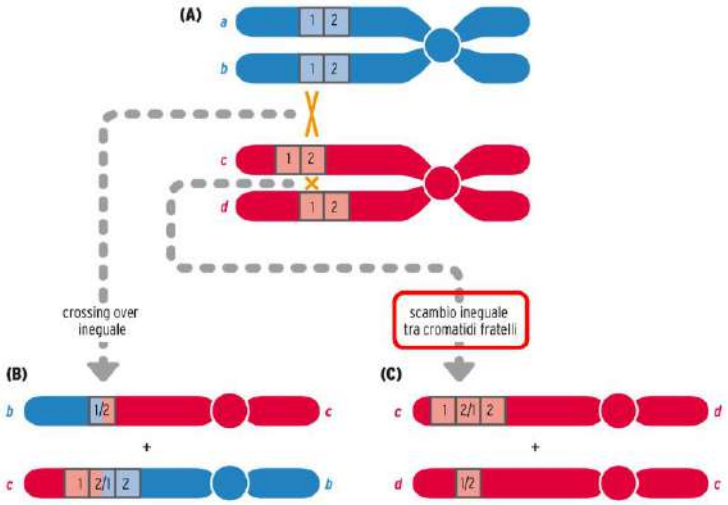
Nell'embrione (fino a 6 settimane) si ha una combinazione $\zeta_2\varepsilon_2$ di globine, nel feto vengono sintetizzate due catene α e due catene γ. Appena prima della nascita si passa dall'espressione della γ-globina a quella della β-globina. Nell'adulto la maggior parte degli invididui è $\alpha_2\beta_2$, una residua percentuale è $\alpha_2\delta_2$.



La presenza di sequenze ripetute può determinare il verificarsi di mutazioni prevalentemente attraverso errori nella ricombinazione omologa. La presenza di geni altamente omologhi nella stessa regione, o di pseudogeni localizzati in prossimità del gene funzionale, può determinare errori di appaiamento ed essere responsabile di crossing-over ineguali, con delezioni su un cromosoma e duplicazioni sull'omologo.

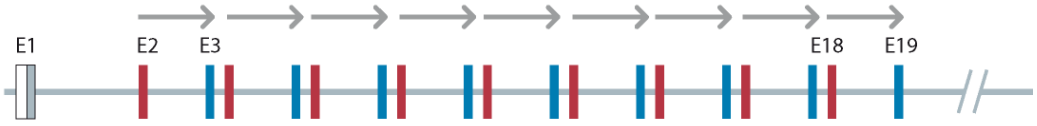


Le duplicazioni geniche possono avvenire anche in seguito a scambio ineguale tra cromatidi fratelli.



I fenomeni di duplicazione possono anche essere intragenici, cioè sullo stesso cromosoma si attivano una serie di ripetizioni, in numero altamente variabile, di una precisa sequenza, che può coinvolgere esoni e domini. Avviene nel 10% dei geni umani; un caso di questo meccansimo è il gene LPA (lipoproteina A). Si può osservare che le sequenze dell'esone 2 e dell'esone 3 si ripetono identiche per ben 8 volte, per l'intera lunghezza del gene codificante.

(A)




```
E2  ------ 17  AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
     E3 ----  AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  130
         131 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  244
         245 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  358
         359 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  472
         473 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  586
         587 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  700
         701 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  814
         815 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
              AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ  928
E18 ----  929 AAPEQSHVVQDCYHGDGQSYRGTYSTTVTGRTCQAWSSMTPHQHNRTTENYPN
     E19 ---  AGLIMNYCRNPDAVAAPYCYTRDPGVRWEYCNLTQCSDAEGTAVAPPTVTPVPSLEAPSEQ 1042
```

In alternativa esoni e domini possono essere "rimescolati". Avviene una migrazione o una selezione di un segmento genico presente in numero ripetuto in una regione del genoma, che si combina con altre porzioni di geni ripetute, a formare un nuovo gene, con caratteristiche derivate da ciascuno dei "pezzi" che lo compongono. E' il caso di un farmaco fibrinolitico, l'attivatore tissutale del plasminogeno (TPA). Esso è in grado di stimolare la formazione della plasmina, enzima in grado di proteggere l'organismo dalla formazione di eventuali occlusioni all'interno dei vasi sanguigni distruggendo alcune strutture presenti nel sangue come la fibrina (proteina utilizzata nella coagulazione del sangue) e i trombi (masse

solide formate da fibrina, piastrine, globuli bianchi e globuli rossi, che si possono formare nei vasi sanguigni). L'organismo umano non produce direttamente la plasmina, ma un suo precursore inattivo, il plasminogeno, che viene attivato da varie sostanze. E' ottenuto tramite tecniche di DNA ricombinante.

E' formato da 4 esoni codificanti domini strutturali diversi: il primo esone è simile a quelli della fibronectina, proteina che lega la fibrina, il secondo esone codifica per un dominio tipico dei fattori di crescita, il terzo e il quarto codificano per strutture "kringle" (legano i coaguli di fibrina), presenti nel gene per il plasminogeno.

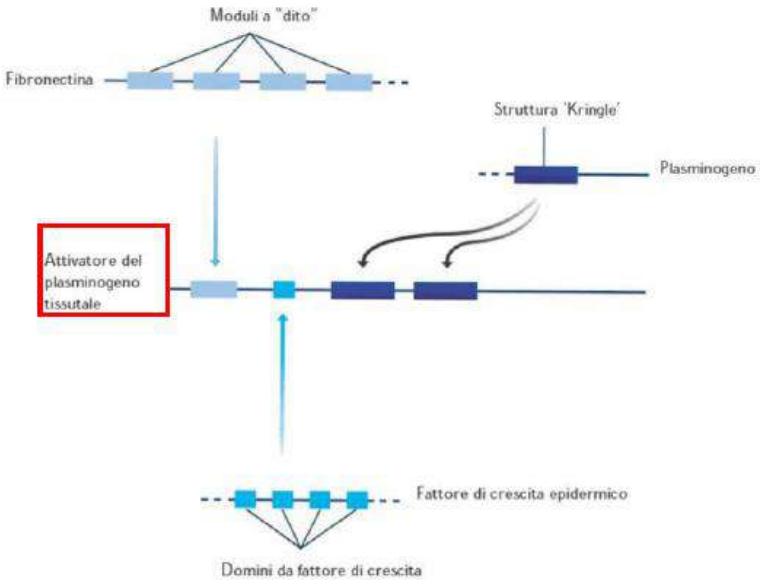


Figura 15.14 La struttura modulare dell'attivatore tissutale del plasminogeno.

Approcci di studio per studiare sequenze di DNA

Per analizzare la popolazione di un DNA complesso è possibile ricorrere a due diversi metodi: l'amplificazione di una sequenza specifica di DNA (clonaggio del DNA) oppure il riconoscimento di una specifica sequenza di DNA tramite ibridazione molecolare. Il clonaggio può essere realizzato in un sistema cellulare o in vitro (PCR).

La scoperta del clonaggio del DNA può probabilmente essere considerata la pietra miliare dell'ingegneria genetica, poiché consente di utilizzare un sistema cellulare per ottenere milioni di copie di una stessa molecola di DNA ricombinante. Quest'ultima si ottiene legando la molecola di DNA di interesse (precedentemente estratta dalle cellule contenenti questa e purificata) a un replicone, una sequenza di DNA capace di replicazione autonoma all'interno di una cellula ospite. Entrambe le molecole di DNA vengono tagliate con endonucleasi di restrizione, che producono sticky ends (estremità adesive) complementari e sono successivamente saldate da una DNA ligasi.

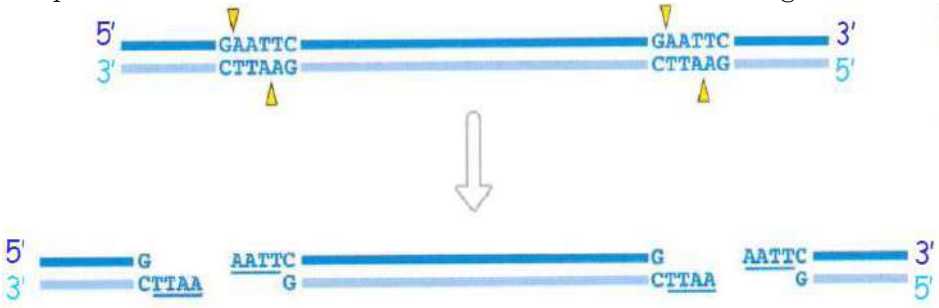


Figura 1 ◀ **Taglio asimmetrico di un DNA a doppio filamento tramite la nucleasi di restrizione *Eco*RI.** Si noti che la sequenza sottolineata AATT è un esempio di estremità 5′ sporgente.

Si ottiene così una molecola circolare di DNA che contiene il replicone, o vettore, e la molecola di interesse, chiamata inserto.

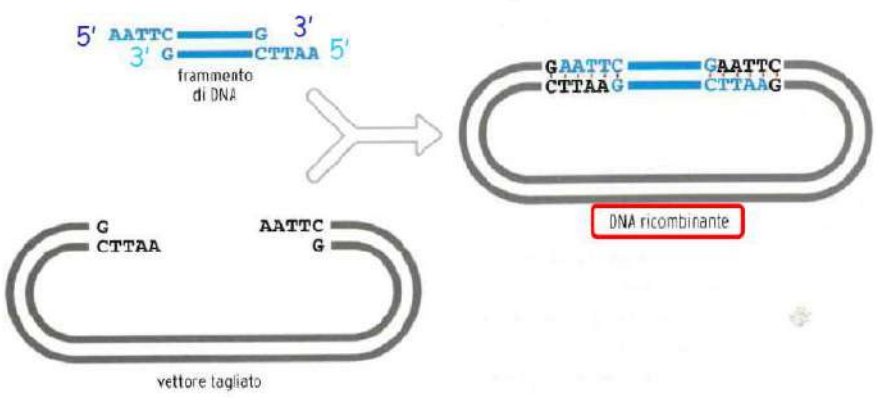


Figura 2 **Formazione di un DNA ricombinante.** In questo esempio il vettore è stato tagliato a livello di un unico sito *Eco*RI per produrre estremità 5' con una sequenza AATT sporgente e il frammento di DNA da clonare ha le stesse sporgenze 5' AATT, essendo stato prodotto da un taglio con *Eco*RI. Le sporgenze AATT sono un esempio di *estremità adesive* perché sono in grado di far formare i legami a idrogeno con altri frammenti che hanno la stessa sequenza sporgente (come si può vedere nel DNA ricombinante; i legami a idrogeno sono rappresentati come linee rosse verticali), facilitando così le interazioni intermolecolari.

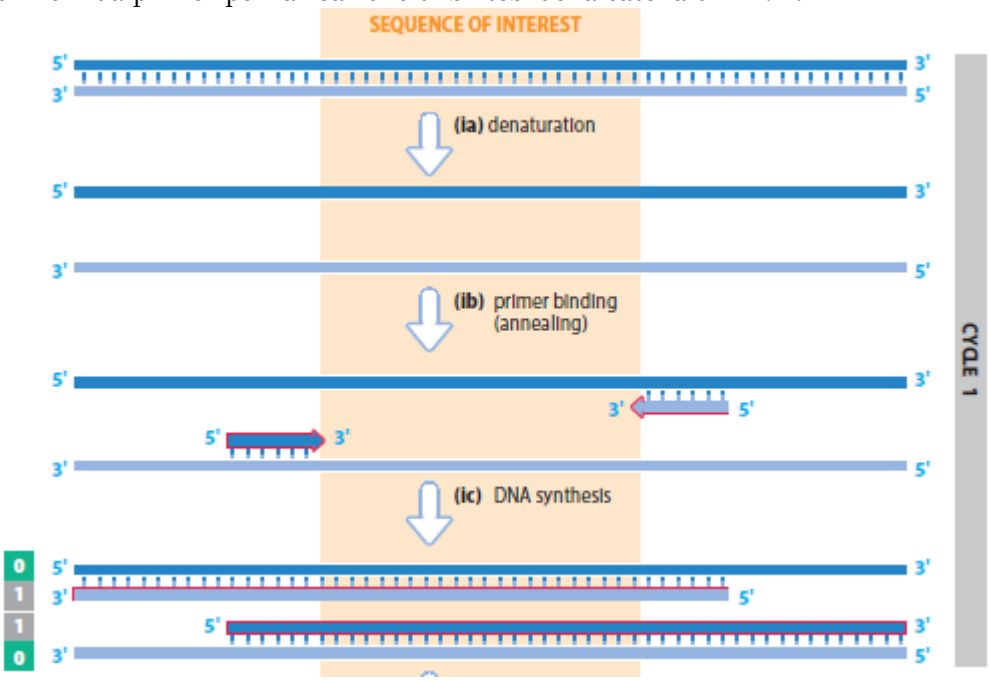
L'unione quindi del DNA omogeneo con il DNA eterogeneo bersaglio produce un DNA ricombinante.
Per il clonaggio vengono applicati alcuni concetti fondamentali della biologia dei procarioti, in particolare la trasformazione, cioè la possibilità di inserire piccole molecole circolari di DNA, dette plasmidi, all'interno della cellula batterica e di ottenerne l'amplificazione in un numero elevato di copie. Il plasmide deve possedere tre requisiti fondamentali: una sequenza ori, alla quale si lega il complesso di duplicazione per replicare il DNA, un marcatore selettivo che permetta di scegliere le sole cellule trasformanti, e un polilinker, la regione del plasmide nella quale può essere inserito un DNA esterno. Il plasmide si replica indipendentemente dal cromosoma dell'ospite. Il limite principale dell'applicazione dei plasmidi è la dimensione dell'inserto, che può raggiungere solo poche chilobasi (5-10). Nel corso degli anni si sono resi disponibili vettori di capacità sempre maggiore, come ad esempio i BAC (cromosomi artificiali batterici), ma anche sistemi cellulari di amplificazione diversi dai procarioti, quali i lieviti YAC (cromosomi artificiali di lievito). La cellula ospite dei BAC è la stessa dei plasmidi: E. coli. La loro massima capacità di clonaggio è di circa 200 kb. Si tratta di vettori plasmidici modificati con un limite rigido sul numero di copie, che consente a grandi frammenti di moltiplicarsi in modo stabile. Gli YAC sono cellule di Saccharomyces cerevisiae (lievito di birra) e possono ospitare inserti lunghi più di 1 Mb. I ricombinanti sono in realtà piccoli cromosomi lineari composti principalmente da DNA umano o da altro DNA estraneo.
Se la cellula ospite proviene da un mammifero, anziché di trasformazione, si parla di transfezione. Nel caso della trasformazione, il DNA plasmidico non viene inserito nel genoma batterico, rimane cioè in forma episomiale, oppure può essere integrato nel genome dell'ospite. Nel primo caso, la molecola di DNA ricombinante si perde nel corso di alcuni cicli di replicazione cellulare. Nel caso dell'integrazione nel genoma dell'ospite, si ha una modificazione permanente del genoma stesso e il DNA esogeno viene trasmesso alle cellule figlie nel corso della mitosi, oppure in maniera mendeliana alla prole nel corso della riproduzione.
La miscela di cellule trasformanti viene sparsa su una piastra di agar contenente antibiotico (per le quali le cellule trasformanti contengono geni di resistenza), che consente alle cellule di crescere e di formare colonie. Ogni colonia batterica sopravvissuta contiene cellule derivate da una sola cellula trasformata, sono cioè cloni identici di questa prima cellula trasformata. Le colonie trasformate vengono poi espanse in brodocolture. Si verifica un'ulteriore crescita, fino ad ottenere quantità molto grandi di

cloni cellulari con sequenze identiche di DNA estraneo. Il DNA ricombinante viene successivamente estratto dalle cellule batteriche.

PCR

La reazione a catena della polimerasi (PCR, Polymerase Chain Reaction) è una tecnica di amplificazione in vitro che consente la sintesi esponenziale di un frammento di DNA a partire da uno stampo di DNA di cui si conosca la sequenza nucleotidica delle regioni terminali (a monte e a valle). Questa tecnica venne messa a punto da Kary B. Mullis e gli valse il Premio Nobel per la Chimica nel 1993. Dagli inizi degli anni '90 si osservò un aumento esponenziale del numero di pubblicazioni basate sull'utilizzo della PCR.

L'amplificazione avviene mediante ripetuti cicli di polimerizzazione che utilizzano come stampo i due filamenti di DNA; fungono da innesco due oligonucleotidi sintetici (primer), che delimitano la regione da amplificare. I due primer sono complementari a sequenze situate sui due filamenti che rappresentano il tratto di DNA bersaglio da amplificare e il loro orientamento in senso 5'-> 3' è in direzione diversa (forward e reverse). Nel caso in cui il campione di partenza sia rappresentato da RNA, la reazione è preceduta da un passaggio di retrotrascrizione dell'RNA a cDNA (DNA copia) utilizzando una DNA polimerasi RNA dipendente detta trascrittasi inversa, un enzima di origine retrovirale che permette di sintetizzare una molecola di DNA usando come stampo un filamento di RNA. Il passaggio di retrotrascrizione a opera della trascrittasi inversa richiede l'utilizzo di una sequenza nucleotidica che funzioni da primer per la reazione di sintesi della catena di DNA.



Per la preparazione della provetta da sottoporre a PCR sono necessari i seguenti elementi:

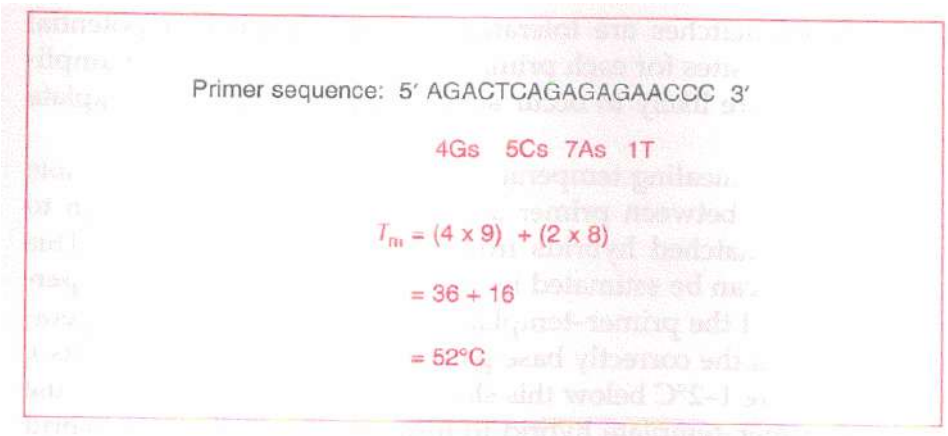
- ✓ Una soluzione buffer. Si utilizza in generale un tampone TRIS-HCl 10-50 mM con un pH compreso tra 8.3 e 9 a 25 °C.
- ✓ Magnesio. Lo ione magnesio Mg^{2+}, utilizzato in concentrazione variabile tra 0,5 e 2,5 mM. Questo parametro deve essere fatto oggetto d'una attenta procedura d'ottimizzazione in quanto può variare anche se si utilizzano diversi primer per amplificare una medesima regione di DNA. La presenza di magnesio condiziona l'attività della polimerasi, l'ibridizzazione dei primer ed aumenta la temperatura cui il DNA stampo si denatura.

- ✓ Desossiribonucleotidi trifosfati. Devono essere in concentrazioni che non variano, per ciascun nucleotide, da 20 a 200 mM.
- ✓ Primers. La lunghezza dei primers è di circa 18 28 nucleotidi. La concentrazione ottimale dei "primers varia da 0,1 a 0,5 mM.
- ✓ DNA polimerasi. Si sceglie di usare una variante termostabile della DNA polimerasi, ricavata dall'organismo Thermus aquaticus, detta Taq polimerasi, che non si disattiva alle elevate temperature a cui viene portato il DNA per denaturarsi nuovamente. La concentrazione di questo enzima può oscillare tra 1 e 5 unità per 100 microlitri di reazione. Esplica la propria attività polimerasica in direzione 5'->3' e aggiunge 50-60 nucleotidi al secondo. E' dotata anche di attività esonucleasica 3'->5', che le permette di sostituire eventuali nucleotidi non correttamente appaiati ("correzione di bozze").
- ✓ Campione iniziale. Può essere costituito da DNA genomico, da RNA, da plasmidi/cosmidi/cromosomi artificiali, campioni patologici, forensi e archeologici.

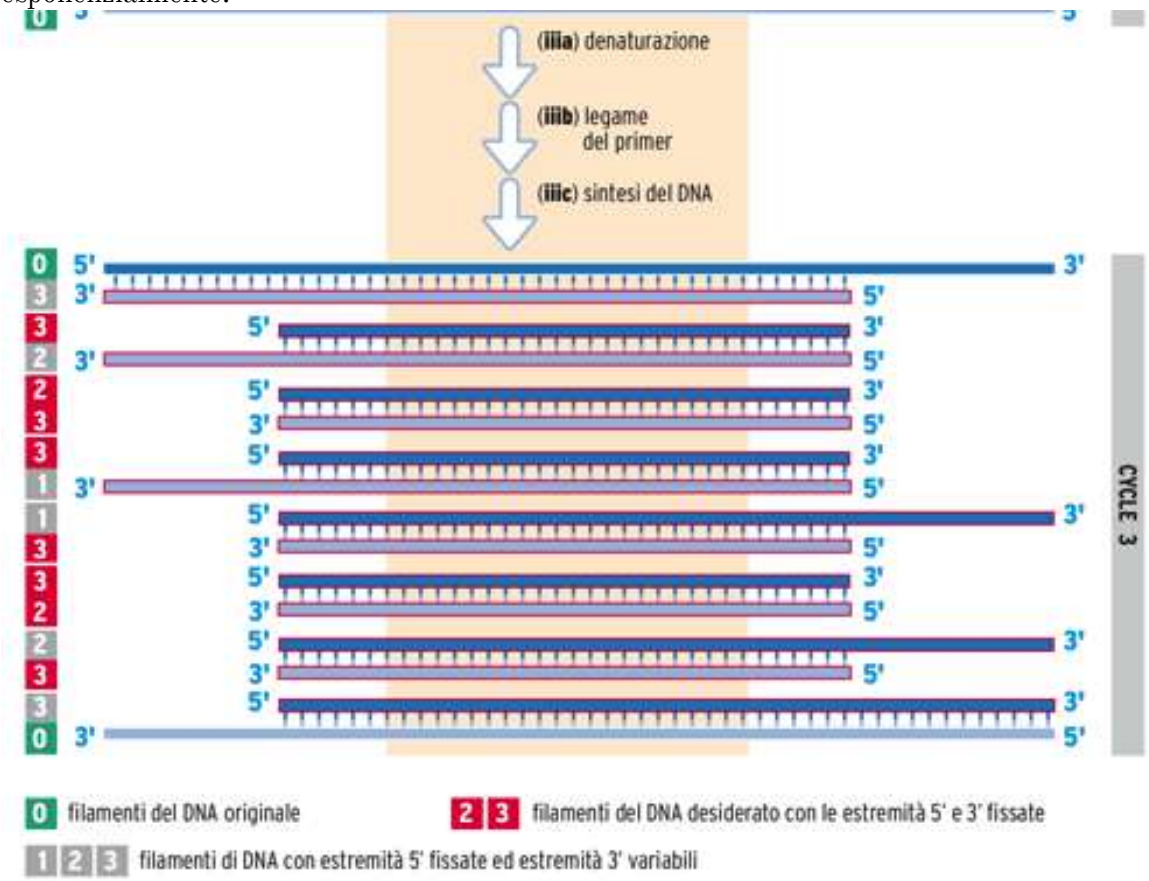
Il requisito essenziale per la PCR è conoscere, almeno in parte, la sequenza nucleotidica a monte e a valle del frammento della regione di DNA di interesse.

1. La soluzione di DNA da replicare, desossiribonucleotidi trifosfati, ioni magnesio, primer e TAQ polimerasi viene portata a una temperatura compresa tra 94 e 99 °C, per 5 minuti. Ci si trova, di conseguenza, in una situazione in cui la doppia elica del DNA viene completamente scissa ed i due filamenti di cui essa è composta sono liberi (fase di denaturazione).
2. Successivamente la temperatura viene abbassata fino a 30-65 °C, per 30 secondi circa, al fine di permettere il legame dei primer alle regioni loro complementari dei filamenti di DNA denaturati (fase di annealing).
3. Infine la temperatura viene alzata fino a 65-75 °C al fine di massimizzare l'azione della TAQ polimerasi che determina un allungamento dei primer legati, utilizzando come stampo il filamento singolo di DNA (fase di prolungamento).
4. La soluzione attraversa nuovamente un ciclo di denaturazione a 94 °C per 30 secondi e si ricomincia dal punto 2.

Ma qual è la temperatura ottimale perché avvenga l'appaiamento tra i primer e la sequenza da amplificare? La temperatura di annealing, così viene definita, per la PCR viene calcolata 2-5 ° C sotto la temperatura di melting dei due primer, ovvero la temperatura alla quale i primer si "sciolgono".
La formula per calcolare la temperatura di melting di una sequenza nucleotidica è la seguente:
$Tm = (4 \text{ X } [G+C]) + (2 \text{ X } [A+T])\,^{\circ}C$.



E' molto importante che i due primer vengano disegnati in modo da avere la stessa Tm.
Se la temperatura di annealing è troppo alta, i primer e il DNA templato rimangono dissociati, mentre se è troppo bassa, si forma un ibrido portatore di uno o più mismatch. L'appaiamento non avviene in maniera precisa, il primer si attacca aspecificamente e gli appaiamenti delle basi primer-bersaglio non sono tutti corretti.
Il ciclo descritto viene ripetuto generalmente per circa 30-40 volte.
Il numero di copie che si forma è così composto: due filamenti sono quelli denaturati nella primissima fase di PCR e si mantengono tali per tutto il corso della reazione; i filamenti con estremità 5' fissa e 3' variabile aumentano in maniera lineare; i filamenti con entrambe le estremità fissate aumentano esponenzialmente.

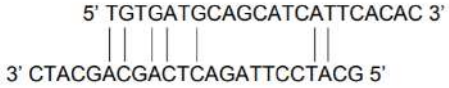


L'immagine raffigura il terzo ciclo di una PCR. I filamenti del DNA originale sono sempre 2, i filamenti con la 3' variabile sono 2 x n (dove n è il numero del ciclo), quindi, in questo caso, 6 e i filamenti con le estremità fissate sono 2^n, in questa circostanza, essendo al terzo ciclo, 8.
La PCR si compone di una fase di ritardo, quando la reazione è appena iniziata, di una fase esponenziale, dove si ottiene un elevatissimo numero di copie del frammento di interesse, e una fase finale di saturazione, dove l'amplificazione, sostanzialmente, si ferma. In genere non si superano i 50 cicli in quanto ad un certo punto la quota di DNA ottenuto raggiunge un plateau (un progressivo rallentamento fino ad un arresto). Ciò avviene, ad esempio, per carenza degli oligonucleotidi usati come inneschi o per diminuzione dei dNTP. Bisogna inoltre considerare che si potrebbe amplificare in maniera eccessiva anche eventuale materiale genomico contaminante. Per allestire una reazione di PCR è necessario disporre di strumenti come un termociclatore, che è una sorta di fornetto in grado di tenere sotto controllo le temperature dei passaggi dei vari cicli, in cui si dispongono dei campioni, solitamente posti in una pluripiastra a 96 pozzetti.

Figura 1.12. Struttura ad hairpin di un possibile primer.

Figura 1.13. Struttura di un self dimer ottenuto dall'appaiamento tra due molecole dello stesso primer.

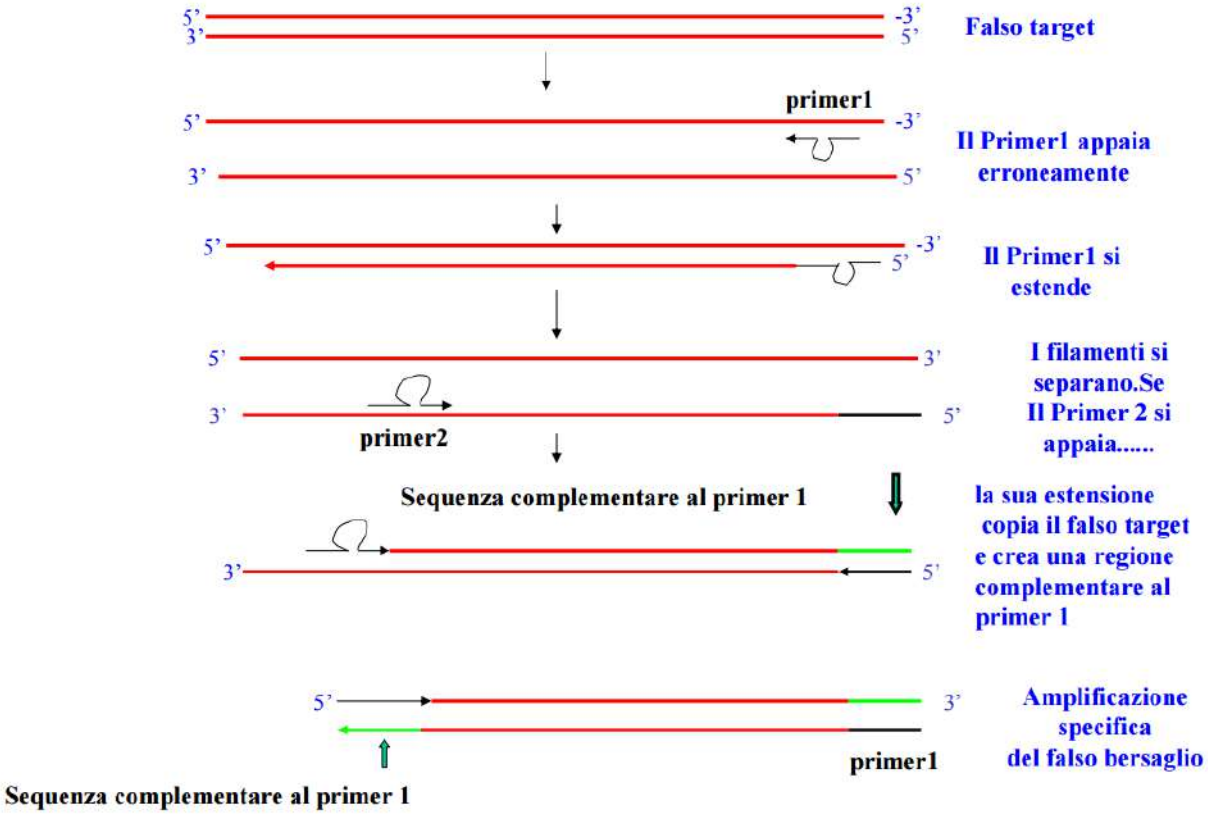
Figura 1.14. Struttura di un cross dimer ottenuto dall'appaiamento tra due molecole di primer diversi.



Questi sono alcuni aspetti da tenere in conto quando si progettano i primer:

- ✓ La lunghezza. Di solito si aggira intorno ai 20 nucleotidi quando il target è una sequenza appartenente ad un DNA genomico complesso. Possono essere più corti se il DNA bersaglio è meno complesso. Una sequenza di 16 bp sarà statisticamente presente una volta ogni 4^{16} (circa 4 miliardi di basi), corrispondenti all'incirca alla grandezza del genoma umano. Più lungo è il primer e piu è stabile, ma è anche maggiore la probabilità che che formi dimeri.
- ✓ La composizione in basi. Sostanziali ripetizioni in tandem di uno o più nucleotidi devono essere evitate. La percentuale totale e la lunghezza in GC devono essere scelti in modo tale che la temperatura di melting dei due primer sia uguale o pressochè identica. Se la temperatura di melting (e, di conseguenza, di annealing) dei due primers è molto diversa si possono verificare amplificazioni asimmetriche o a singolo filamento.
- ✓ La struttura secondaria. Bisogna evitare sequenze tendenti a strutture secondarie che potrebbero formare forcine o altri complessi.
- ✓ L'estremità 3'. La complementarità nucleotidica delle due basi all'estremità 3' di uno dei due primer devono essere evitate. Altrimenti si verifica la formazione di dimeri, che riduce l'efficacia di amplificazione.

Se i primer sono aspecifici, cioè sono in grado di legare altre sequenze, oltre a quella di nostro interesse, anche ammettendo qualche malappaiamento, c'è il rischio concreto che venga amplificato il segmento genico sbagliato, rendendo inutile l'intera reazione.



Il procedimento erroneo si verifica così:

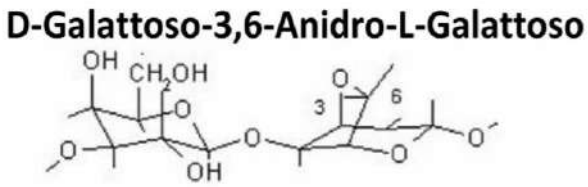
1. Il primer 1(forward o reverse) si appaia erroneamente ad una regione dove la complementarità non è perfetta, quindi forzando uno o più mismatch, legandosi in maniera stabile.
2. Il primer 1 si estende, terminando la trascrizione del proprio filamento complementare a quello di partenza. In questa fase di allungamento l'appaiamento è perfetto, poiché la DNA polimerasi aggiunge i desossinucleotidi complementari al filamento del campione.
3. Avviene la denaturazione dei due filamenti. In questa fase entra in gioco il primer 2, che ha direzione opposta rispetto al primer 1. Il primer 2 si appaia al filamento sintetizzato dal primer 1, anche con una complementarità non totale. Dalla regione di legame primer-filamento il primer 2 prosegue la sintesi del nuovo filamento, che, nella porzione terminale, corrisponderà perfettamente alla sequenza complementare del primer 1. Il falso target viene copiato e il primer 1, d'ora in avanti, avrà una regione complementare a sé, alla quale legarsi solidamente per i prossimi cicli.
4. Al ciclo successivo il primer 1 trascrive il filamento sintetizzato nel ciclo precedente dal primer 2, trascrivendo il sito dove prima il primer 2 si era legato forzatamente e offrendo adesso, per tutti i cicli successivi, una regione di perfetta complementarità per il primer 2.
5. In questo modo si verifica l'amplificazione specifica del falso bersaglio.

L'elettroforesi

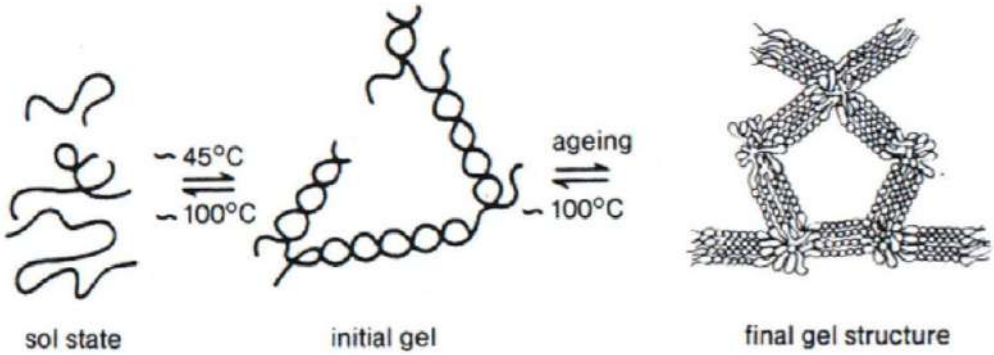
L'elettroforesi è una tecnica che utilizza un campo elettrico per separare e visualizzare frammenti di DNA in base al loro peso molecolare ed alla loro carica. Gli acidi nucleici, migrano sempre verso il polo positivo per la presenza, nella loro molecola, della carica negativa data dai gruppi fosfato.

Il supporto utilizzato per eseguire l'elettroforesi è il gel, che si ottiene dalla fusione e successiva gelificazione di una soluzione contenente agarosio e, come solvente, TAE.

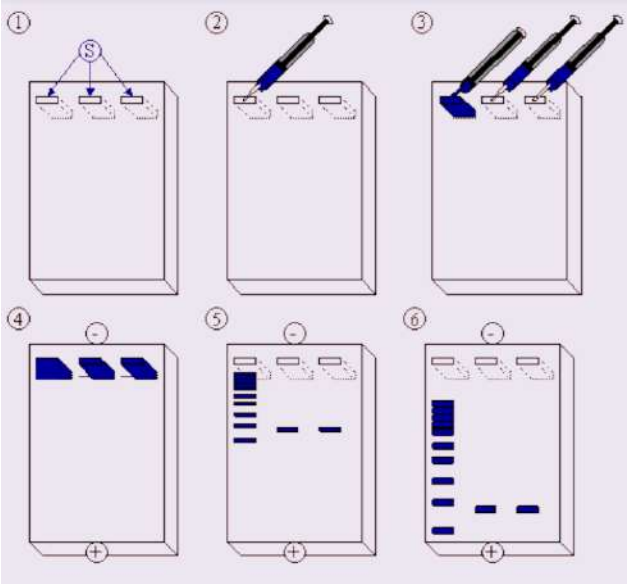
L'elettroforesi in gel è il metodo standard utilizzato per separare, identificare e purificare frammenti di DNA. L'agarosio è un polimero di carboidrati estratto dalle alghe. Fuso e gelificato, forma una matrice, la cui porosità dipende dalla concentrazione. Il risultato è che i frammenti di DNA si separeranno nella matrice in funzione del peso molecolare e della forma Elevate concentrazioni si usano per separare frammenti più piccoli, concentrazioni più basse frammenti più grandi.
La concentrazione dell'agarosio può essere variata per ottimizzare la migrazione. 0,8% riesce a separare un range di dimensioni da circa 500 bp a circa 20.000 bp. 2%, si osserverà una migliore separazione dei piccoli frammenti, perché anche questi saranno rallentati dalla matrice più concentrata.



L'agarosio è un polimero lineare che forma una matrice semisolida avente pori di dimensione diversa in funzione della concentrazione utilizzata. La sua concentrazione determina il potere risolutivo del gel.



Le soluzioni hanno lo scopo, di determinare e mantenere stabili il pH e la concentrazione di ioni del gel. TAE e TBE sono i tamponi più comunemente usati per l'elettroforesi.
T sta per Tris/HCl, che consente il mantenimento di un valore costante di pH della soluzione. E sta per EDTA, che chela i cationi bivalenti, come il magnesio. Questo è importante, perché la maggior parte delle nucleasi richiede cationi bivalenti per funzionare. A o B sta per acido Borico o Acetico che forniscono l'appropriata forza ionica al tampone.

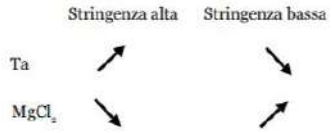


Ogni pozzetto del gel viene caricato con 5 µl di ciascun estratto mescolato con 1 µl di un "gel loading buffer" contenente 0,25% xylene.
Il buffer serve per tre scopi: • aumenta la densità del campione permettendo al DNA di penetrare in maniera uniforme nei pozzetti del gel;
•colora il campione semplificando in tal modo il caricamento;
•contiene dei coloranti che si muovono in un campo elettrico verso l'anodo con una velocità prevedibile.
Come in tutte le tecniche laboratoristiche è necessario allestire sia un C+ che un C- per evidenziare eventuali falsi positivi.

Il controllo positivo è costituito da una provetta nella quale si amplifica un'altra porzione genomica, che può trovarsi su un altro cromosoma rispetto al segmento genico di interesse. L'utilità di questa procedura è di verificare che il DNA di cui stiamo cercando l'amplificazione non sia frammentato (se il controllo positivo ha successo, vuol dire che il cromosoma del controllo positivo è intatto, da cui possiamo plausibilmente dedurre che l'intero genoma è integro), quindi si evita il rischio di un'amplificazione aspecifica. Se il DNA fosse spezzettato i primer non si legherebbero al DNA target, ma si attaccherebbero ad un segmento random tra i tanti presenti, anche con mismatch.
Nell'allestimento del controllo negativo (o "bianco") vanno messi tutti i componenti della miscela di PCR tranne il DNA bersaglio. Se la corsia in elettroforesi del controllo negativo è priva di ogni bandeggio, significa che le operazioni di PCR sono state effettuate in condizioni totali di igiene, dopo aver adottato le procedure di sicurezza e di anticontaminazione. Se invece il controllo negativo mostra una contaminazione, i casi sono due: la strumentazione usata per lavorare era già compromessa prima dell'effettivo utilizzo oppure gli operatori non sono stati adeguatamente attenti nel manipolare le sostanze e le provette, provocando l'"inquinamento" dei campioni analizzati.

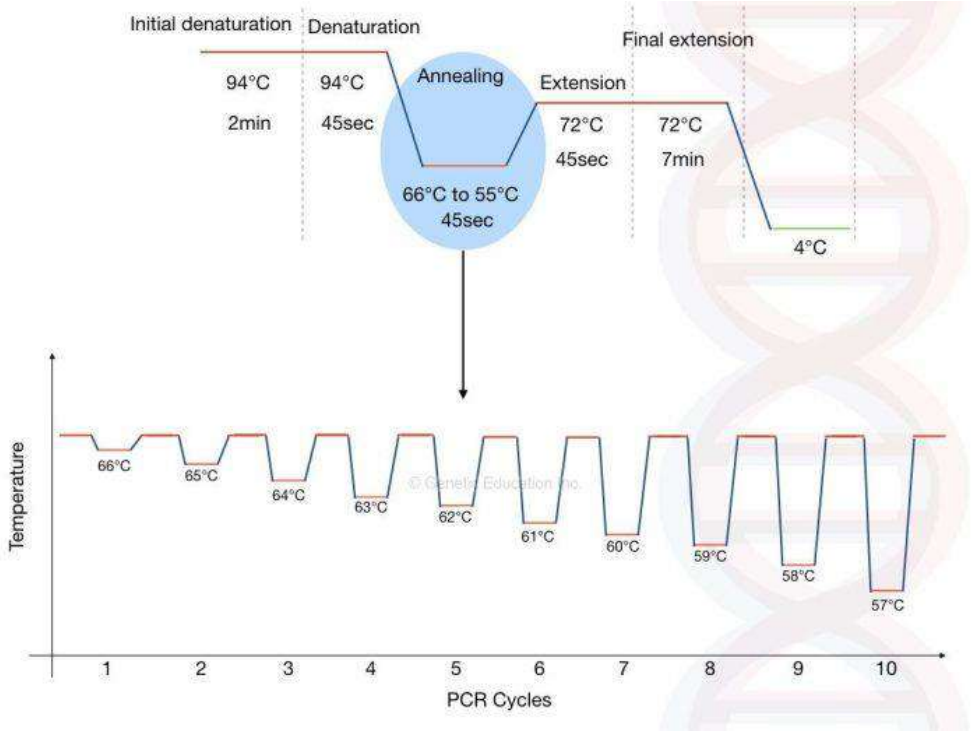
Strategie per aumentare la specificità della PCR

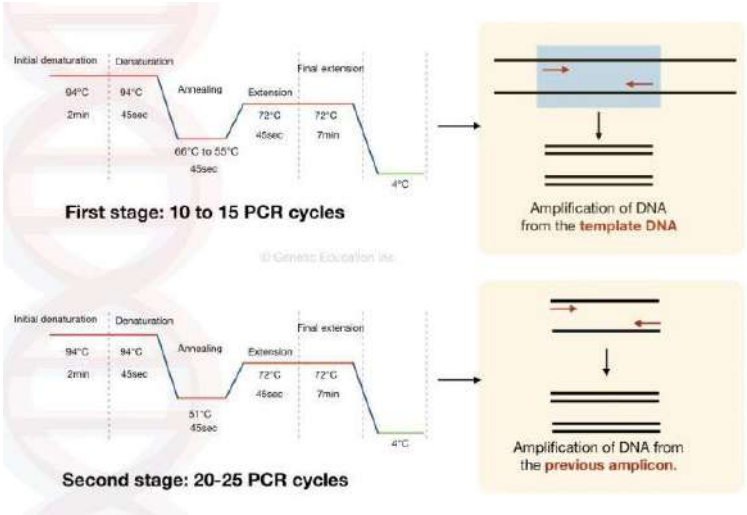
Figura 1.4. Effetto del cloruro di magnesio (MgCl2) e della temperatura di annealing (Ta) sulla stringenza (specificità) di una reazione di PCR. Le frecce in alto e in basso indicano rispettivamente un aumento o una diminuzione della Ta e del MgCl2.



- ✓ Nel caso si debba mettere a punto una PCR è di norma preferibile avere a disposizione il tampone di reazione privo di $MgCl_2$, ed aggiungere il sale separatamente. Infatti, la concentrazione degli ioni ha due ruoli: da un lato è un cofattore essenziale per la Taq DNA polimerasi per il caricamento dei nucleotidi, dall'altro la concentrazione di $Mg2^+$ influenza l'appaiamento dei primer allo stampo. Maggiore è la concentrazione di $MgCl_2$ minore è la specificità dell'appaiamento. La variazione delle concentrazioni di $MgCl_2$ gioca quindi un ruolo simile (ma opposto) a quello della temperatura di annealing e spesso la messa a punto di un'elevata (o bassa) stringenza (specificità) nella PCR è basata tanto sulla Ta quanto sulla concentrazione di $MgCl_2$.
- ✓ La PCR Hot Start è una tecnica che riduce l'amplificazione aspecifica e offre il vantaggio di permettere l'allestimento della reazione a temperatura ambiente. Le polimerasi utilizzate nelle PCR Hot Start non sono reattive a temperatura ambiente. L'attività delle polimerasi a queste temperature può essere inibita attraverso diversi meccanismi, tra cui l'interazione con anticorpi, modificazioni chimiche, e la tecnologia degli aptameri (acido nucleico sintetico di circa 70-80 nucleotidi in grado di riconoscere e legarsi ad una grande varietà di molecole). Alle temperature a cui può avvenire la reazione raggiunte durante i cicli di PCR, la polimerasi si dissocia dal proprio inibitore e intraprende la polimerizzazione. L'utilizzo di DNA polimerasi Hot Start è raccomandata soprattutto per procedimenti di screening ad alto rendimento, esperimenti che richiedono un elevato grado di specificità, o anche PCR di routine per le quali si renda preferibile un grado di sicurezza ulteriore fornito da un enzima Hot Start.
- ✓ La PCR touchdown è una variante della PCR convenzionale nella quale l'elevata specificità di amplificazione è ottenuta attraverso la riduzione dell'amplificazione di falsi target con una progressiva riduzione della temperatura di annealing dopo ogni ciclo di PCR.

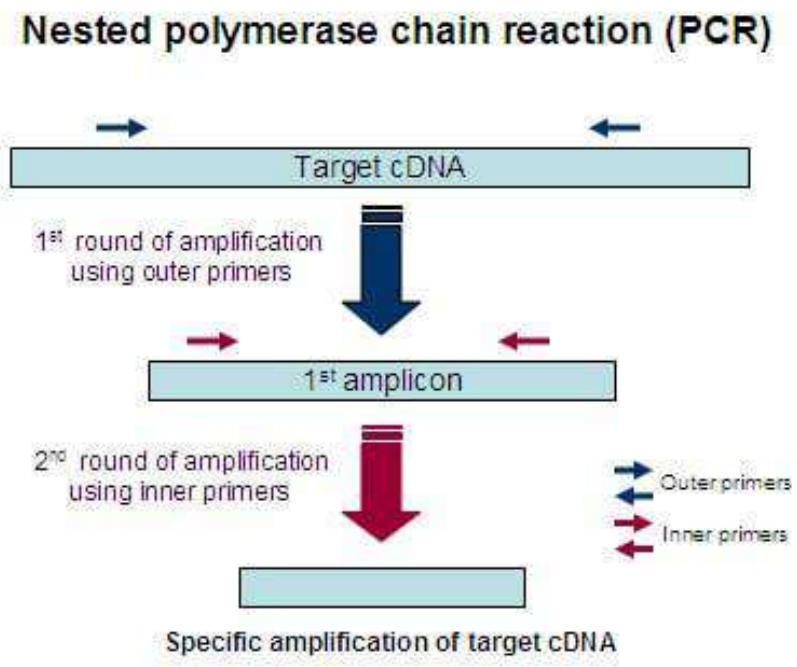
Nella PCR touchdown, la temperatura di annealing è selezionata 10 °C più alta della temperatura di melting. Perciò, i primer possono legarsi soltanto alla propria sequenza specifica e complementare. Dopo ogni ciclo, la temperatura viene abbassata di 1 °C. Anche se il DNA non viene amplificato nei cicli iniziali a causa di una temperatura troppo alta, verrà comunque amplificato in uno dei cicli successivi, dal momento che la temperatura diminuisce progressivamente di ciclo in ciclo. L'aspetto più vantaggioso di questa tecnica è che, nonostante il progressivo abbassamento della temperatura, non si verifica la formazione di legami non specifici dopo l'amplificazione del primo templato. Una volta che il DNA di interesse è stato amplificato con successo, quello stesso DNA amplificato funge da templato per i successivi cicli di PCR.





Se la temperatura di annealing di questa PCR è 55 °C. Bisogna dividere i cicli di PCR in due fasi. Nella prima fase si abbassa la temperatura di annealing di 1 °C dopo ogni ciclo. Se la T_a è 55 °C, scegliamo 66 °C di temperatura per il primo ciclo di amplificazione. Si predispongono 15 cicli, quindi la temperatura, al termine di questa prima fase, si abbasserà di 15 °C. La temperatura di annealing raggiunta sarà di 51 °C. Nella seconda fase si mantiene questa temperatura di annealing sarà mantenuta per il numero restante di cicli. La combinazione della PCR touchdown con la PCR hot start rende questa tecnica ancora più avanzata.

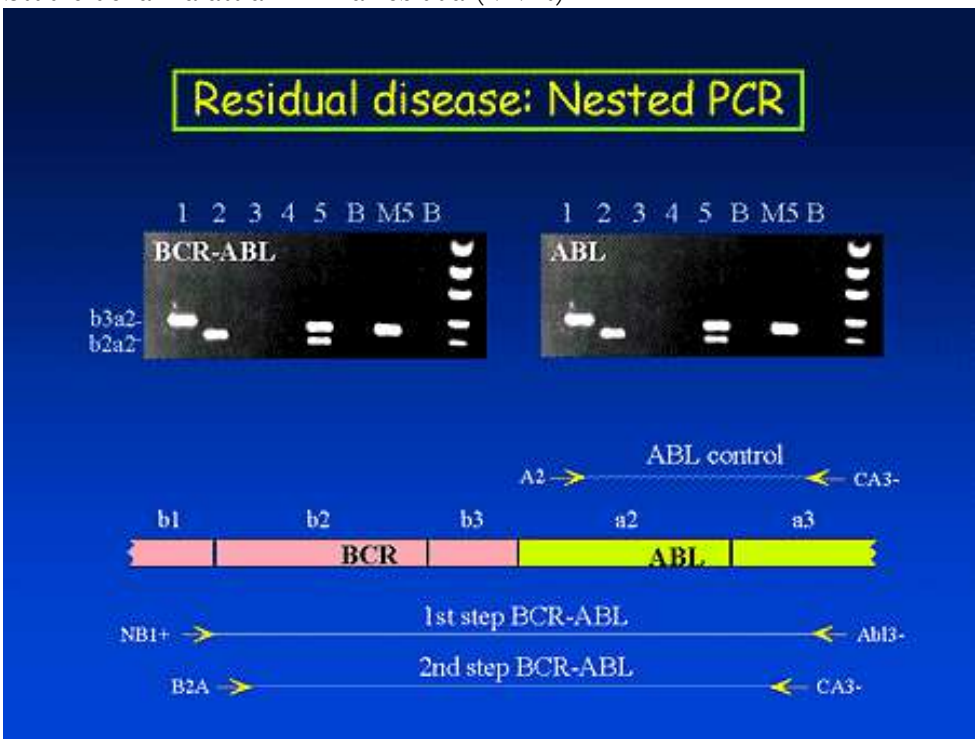
- ✓ La nested PCR è una variante della tecnica di PCR, utilizzata per aumentare la specificità dell'amplificazione, che prevede due distinte e successive reazioni di amplificazione:
  Nella prima amplificazione (prima PCR) si utilizza una coppia di primers più esterni al frammento di DNA da amplificare.
  2. Nella seconda amplificazione (seconda PCR) si utilizza una coppia di primers che amplificano un frammento interno a quello amplificato nella prima reazione. In pratica, una frazione del prodotto amplificato nella prima PCR viene trasferito in una seconda provetta contenente una nuova miscela di reazione, uguale alla precedente per tutti i componenti ad eccezione dei primers. Il target dell'amplificazione è presente in condizioni completamente diverse nelle due reazioni di PCR. Infatti, nella prima ci sono poche copie di sequenza bersaglio in un "background" di DNA complesso e abbondante mentre nella seconda PCR il numero di sequenze bersaglio è maggiore e il "background" è significativamente ridotto.
  Si tratta, quindi, di una modifica della PCR tradizionale destinata a ridurre o a eliminare eventuali prodotti aspecifici non desiderati in seguito all'amplificazione di siti di legame di primer inaspettati (primers che possono legarsi a regioni di DNA non corrette) .
  In questo modo, si viene ad aumentare la resa e la specificità di PCR particolarmente difficili.
  Se il prodotto di amplificazione fosse aspecifico la seconda PCR non andrebbe a buon fine.
  Questa tecnica trova però sempre minore applicazione in ragione degli alti rischi di contaminazione tra i diversi campioni per il fatto che le provette contenenti i prodotti di amplificazione devono necessariamente essere manipolate nella stessa area in cui sono presenti le miscele della seconda reazione.



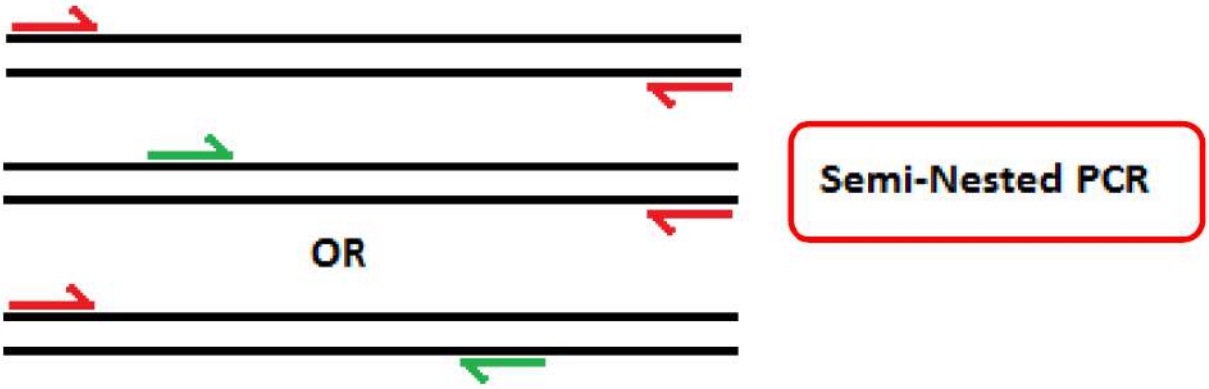
Per ridurre questo rischio si consiglia di minimizzare il numero di cicli della prima PCR. Inoltre, nella seconda reazione viene prodotta una quantità elevata di frammenti bersaglio che rappresenta la più pericolosa fonte di contaminazione di "carry – over".
Questa tecnica si applica quando si tratta di amplificare:

- ✓ Campioni frammentati, contaminati o poco concentrati;
- ✓ Tracce di sangue conservato in eparina (blocca la processività della polimerasi);
- ✓ Campioni storici;

- ✓ Campioni mantenuti in paraffina (provoca la frammentazione del DNA);
- ✓ Studio della malattia minima residua (MMR)



- ✓ La hemi-nested (o semi-nested) funziona analogamente, ma nella seconda PCR utilizzi uno dei due primers della prima associato con un solo primer più interno.



Quali problematiche si possono presentare nel corso della PCR e dell'elettroforesi e a cosa sono dovute?
Se non si osserva nessun amplicone o scarse tracce, la causa può essere una concentrazione scorretta dei reagenti, una progettazione sbagliata dei primer, la presenza di agenti inibitori, una temperatura di annealing troppo alta, un numero di cicli di PCR insufficiente, un malfunzionamento del termociclatore.
Se le bande evidenziate sono aspecifiche, si è verificato, probabilmente, un numero troppo elevato di cicli o una concentrazione troppo alta di reagenti, la temperatura di annealing è troppo bassa, il disegno dei primer può essere erroneo.
Se la PCR ha amplificato solo alcuni campioni bisogna assicurarsi che il DNA templato fosse stato dosato correttamente o si è in presenza di polimorfismi inaspettati che ostacolano l'annealing di un primer.
Se il prodotto ha dimensioni errate significa che l'appaiamento di primer è avvenuto in regioni non corrette (questo avviene frequentemente nei casi di famiglie geniche, pseudogeni, sequenze ripetute).

Gli SNP (Single Nucleotide Polymorphisms)

Circa ogni 1000 basi fisse del DNA individuale esiste una possibile variazione e l'elemento che si modifica viene chiamato "snp" (single nucleotide polymorphism).

Il 99.9% del genoma è identico per tutti mentre lo 0.1% è il solo responsabile delle variazioni e della specificità di ogni individuo. Gli SNP attualmente censiti nell'apposita banca del National Centre for Biotechnology Information (NCBI) (dbSNP) sono oltre 150 milioni e rappresentano in genere varianti di sequenza limitate a un solo nucleotide che è sostituito da un altro in una percentuale della popolazione convenzionalmente superiore all'1% (se è sotto l'1% si considera una mutazione).

Questi polimorfismi hanno un ruolo nell'indurre la suscettibilità specifica delle risposte dell'organismo nei confronti degli stimoli interni (endogeni) ed esterni (esogeni). Va sottolineato che, per indicare lo tesso fenomeno, mentre il termine "mutazione" indica un cambiamento deleterio per l'organismo, il più generico "polimorfismo" si riferisce a una variante genetica che può essere anch'essa portatrice di danni per la cellula oppure contribuire soltanto alla variabilità delle popolazioni.

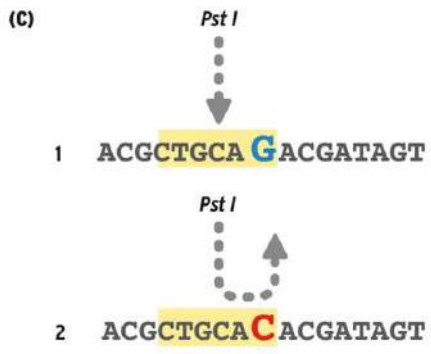
Per cui conoscendo i polimorfismi specifici dell'individuo è possibile, in termini di stima, valutare l'aumento del rischio specifico rispetto alla popolazione generale.

Gli SNP possono influenzare la probabilità di condizionare le risposte dell'organismo in diversi modi, alterando ad esempio l'espressione dei geni e quindi causando, a seconda dei casi, una maggiore o una minore quantità di una determinata proteina rispetto alla norma; oppure possono alterare l'efficienza delle proteine che vengono prodotte, provocando la produzione di proteine che funzionano male o che sono più instabili.

Alcuni di questi SNP possono influenzare la sensibilità a varie intolleranze alimentari (lattosio, glutine) o cutanee, o a predisposizioni a serie patologie come l'osteoporosi, le malattie cardiovascolari, l'Alzheimer, l'obesità. Altri SNP influenzerebbero la capacità dei tessuti ad utilizzare in modo appropriato il glucosio (lo zucchero circolante), con inevitabili conseguenze sul piano del metabolismo individuale. Un altro gruppo contribuisce alla qualità relativa alla produzione di sostanze coinvolte nei processi di difesa dell'organismo verso i danni ossidativi, mediati dai cosiddetti radicali liberi o all'attività infiammatoria con serie conseguenze sulla salute.

Gli SNP sono quindi mutazione puntiformi che possono produrre o abolire siti di riconoscimento per enzimi di restrizione del DNA e, in questo caso, corrispondono ai primi marcatori impiegati negli anni Ottante del secolo scorso per lo studio del genoma, ovvero i RFLP (Restriction Fragment Length Polymorphism).

I RFLP



Il termine RFLP è acronimo dell'inglese Restriction Fragment Length Polymorphism, che significa "polimorfismo della lunghezza dei frammenti di restrizione".

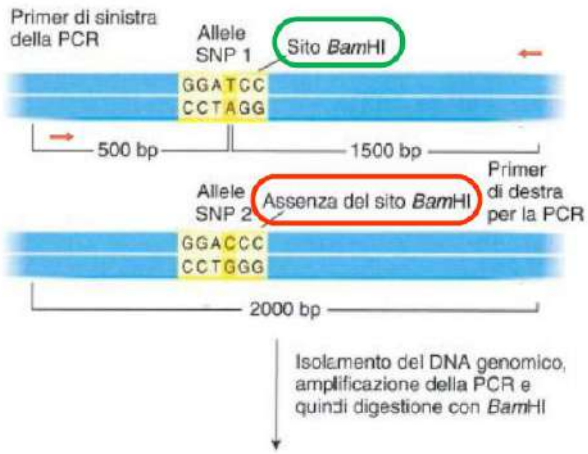
Per inciso, gli enzimi di restrizione permettono di tagliare il DNA in frammenti in corrispondenza di specifiche sequenze dette palindromiche, cioè che si possono leggere in entrambi i sensi sui due filamenti complementari; per esempio l'enzima *BamHI* riconosce la seguente sequenza (lo slash

/ rappresenta il punto di taglio; il colore evidenzia la simmetria nelle sequenze su ciascun filamento e sui due filamenti complementari):

5'... G / G A T C C...3'

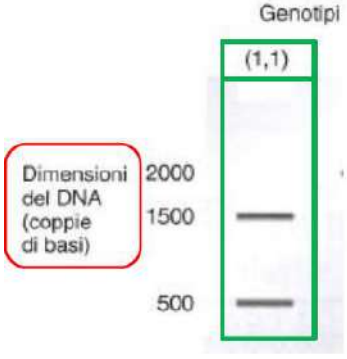
3'... C C T A G / G...5'

In origine la tecnica RFLP fu messa a punto e utilizzata per evidenziare un polimorfismo consistente nella presenza o assenza di un sito di restrizione, dovuto a mutazioni che vanno a colpire il sito di restrizione stesso facendolo scomparire se esiste, o creandone uno nuovo. Successivamente però il termine ha acquisito un significato più ampio, poiché sono stati evidenziati dei tratti di DNA in cui, pur non essendovi mutazioni nei siti di restrizione, vi sono tuttavia polimorfismi di lunghezza dei frammenti. Un'analisi della sequenza di tali frammenti ha messo in luce la presenza di DNA ripetitivo. Attualmente quindi, nel caso di accertamenti di parentela, per RFLP si intende un polimorfismo di lunghezza dovuto a brevi sequenze ripetute in tandem, che essendo molto variabili portano ad un numero così alto di varianti che prendendo due individui dalla stessa popolazione ci si aspetta che questi abbiano la stessa serie di frammenti di restrizione solo se sono consanguinei.

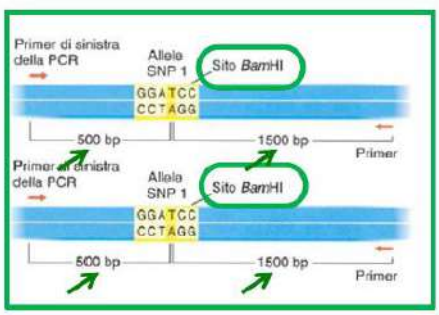


Il metodo della PCR per studiare gli SNP che influenzano i siti di restrizione. Un frammento di 2000 bp del cromosoma ha alleli SNP a 500 bp dall'estremità di sinistra. Il variare della coppia T-A dell'allele SNP 1 (in alto) in G-C nell'allele SNP 2 (in basso) modifica un sito *BamHI* in una sequenza che non viene riconosciuta dall'enzima di restrizione. La PCR dei campioni di DNA derivati da tre individui che hanno genotipi SNP diversi, usando i primer di sinistra e di destra mostrati, seguita da digestione con *BamHI* produce il profilo di bandeggio illustrato di seguito.

INDIVIDUO 1 (OMOZIGOTE PER L'ALLELE 1)

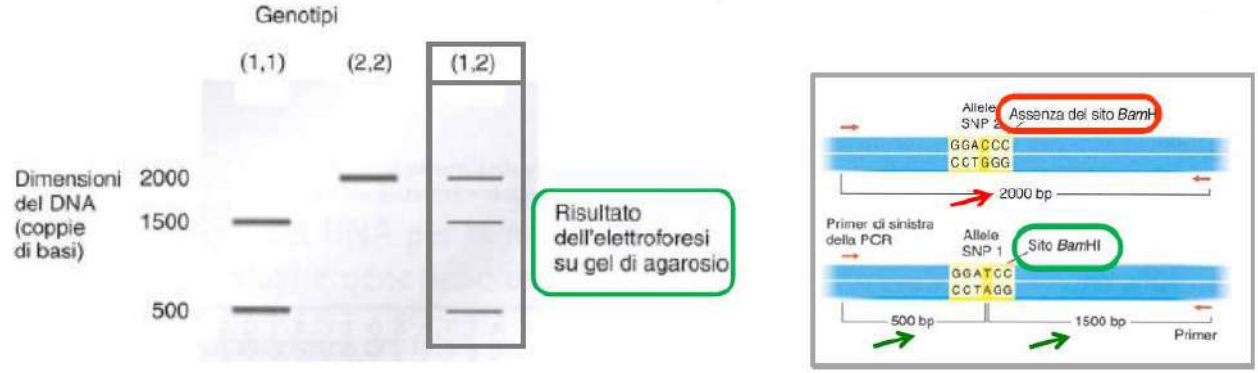


Risultato dell'elettroforesi su gel di agarosio



INDIVIDUO 2 (OMOZIGOTE PER L'ALLELE 2)

INDIVIDUO 3 (ETEROZIGOTE)



In conclusione, se una variazione di sequenza puntiforme crea o abolisce un sito di restrizione, può essere facilmente evidenziata sottoponendo l'amplimero all'azione dell'enzima. La semplice elettroforesi su gel di agarosio o di poliacrilammide, seguita da colorazione con bromuro di etidio, permette di visualizzare direttamente la presenza o l'assenza del taglio enzimatico.

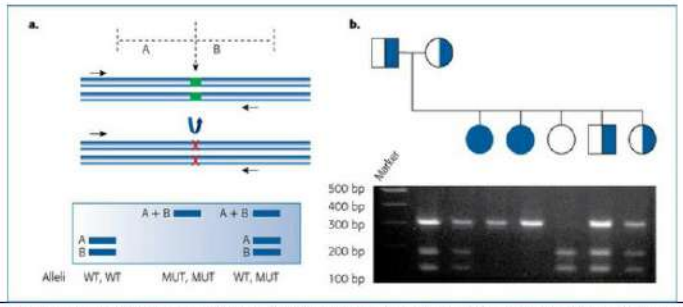
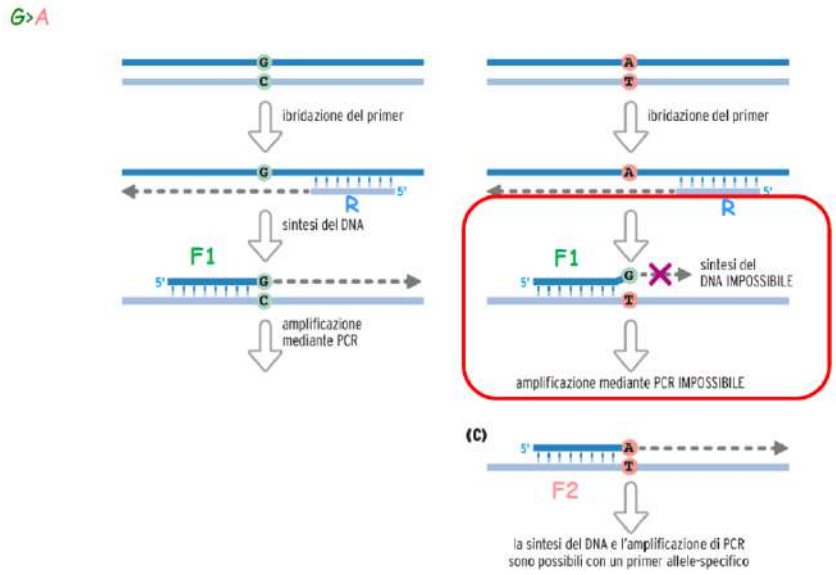


Figura 3.7 Analisi di una mutazione nota mediante digestione con enzima di restrizione. La mutazione abolisce un sito di restrizione. Il prodotto di PCR viene digerito con lenzima appropriato: il frammento corrispondente allallele normale verrà tagliato in due frammenti (A e B), mentre il frammento corrispondente allallele in cui la mutazione ha abolito il sito di restrizione non verrà tagliato (A + B). Dopo elettroforesi su gel di agarosio è possibile discriminare il soggetto omozigote normale (presenza di due bande) dal soggetto omozigote mutato (presenza di una banda) e dal soggetto eterozigote (presenza di 3 bande corrispondenti al frammento intero e a quello digerito) (a). Analisi di restrizione per la valutazione della presenza della mutazione c.191C > T (p.Q33X) nel gene *TREM2* in una famiglia con una forma di demenza precoce a trasmissione autosomica recessiva. La mutazione abolisce il sito di taglio per lenzima *PstI*: nei due soggetti affetti lanalisi di restrizione identifica la presenza di ununica banda elettroforetica di 300 paia di basi, corrispondente al prodotto di amplificazione non digerito, mentre nei soggetti eterozigoti si identificano tre bande elettroforetiche di 300, 200 e 100 paia di basi corrispondenti rispettivamente allallele mutato (assenza del sito di taglio) e allallele normale (b).

ARMS (Amplification Refractory Mutation System)



Questa tecnica di analisi prevede l'amplificazione del DNA tramite PCR utilizzando coppie di primer allele-specifici. Per l'analisi di ogni alterazione vengono sintetizzati tre diversi primer; due primer forward, allele-specifici, che differiscono tra loro per il nucleotide in posizione 3' (complementare o alla sequenza normale o a quella mutata), e un primer reverse comune. Per l'analisi di ogni campione si eseguono due diverse amplificazioni utilizzando, in combinazione con il primer comune, ciascuno dei due primer allele-specifici. La mix 1 conterrà dunque: la Taq polimerasi, $MgCl_2$, acqua, la soluzione buffer, i dNTPs, il DNA, il primer forward 1 (F1) e il primer reverse. La mix 2 si differenzierà da questa prima mix soltanto per la presenza del primer forward 2 (F2), mentre il resto rimarrà invariato.
Il DNA non viene amplificato quando il primer non è perfettamente complementare alla sequenza in esame per tutta la sua lunghezza. La visualizzazione dei prodotti amplificati avviene dopo elettroforesi su gel di agarosio e colorazione con etidio bromuro. La ARMS PCR rappresenta una buona tecnica di screening per la valutazione di un gran numero di campioni per un pannello di mutazioni non molto ampio.

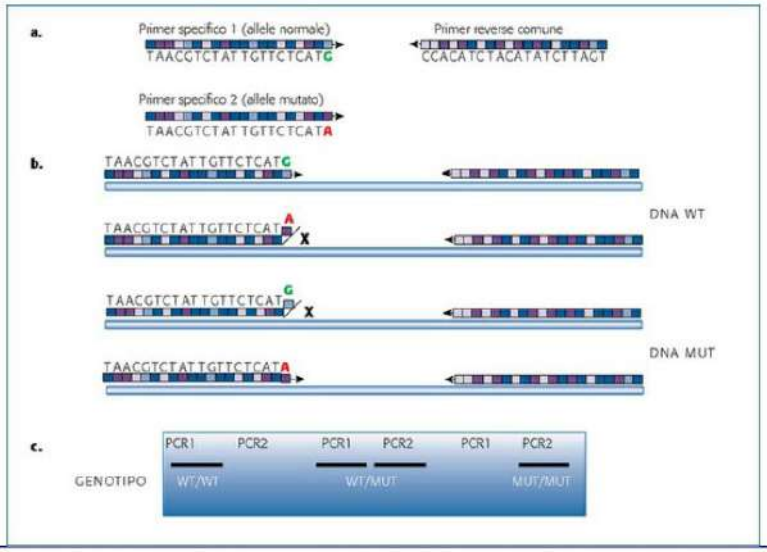
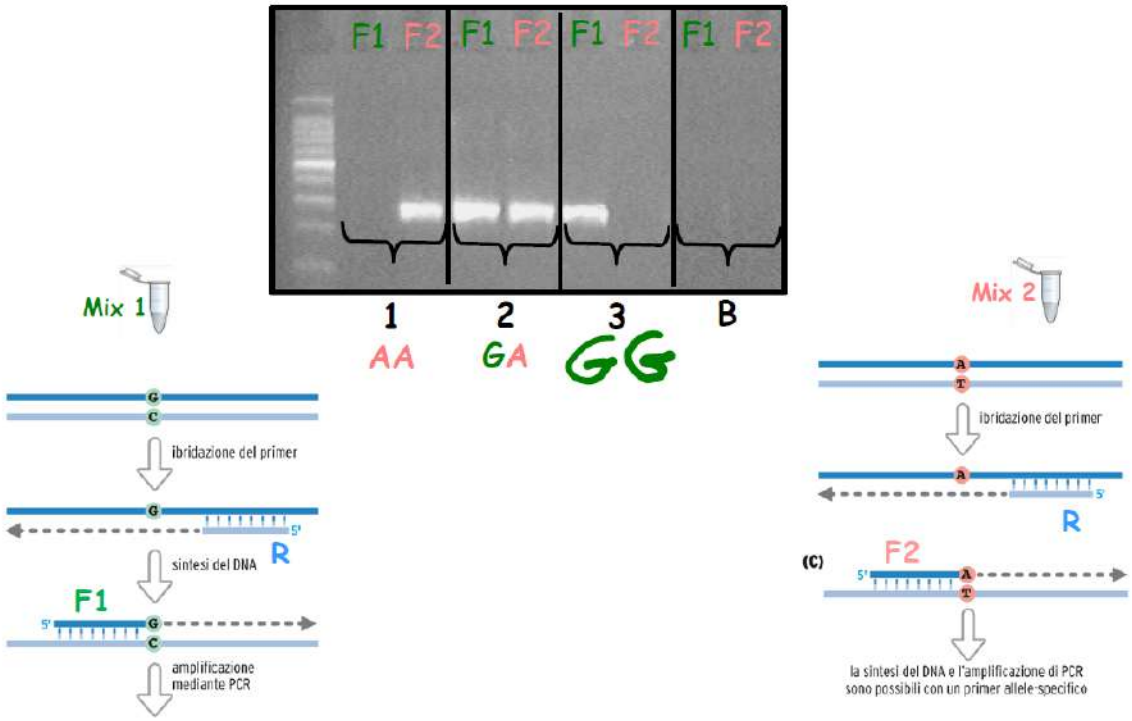
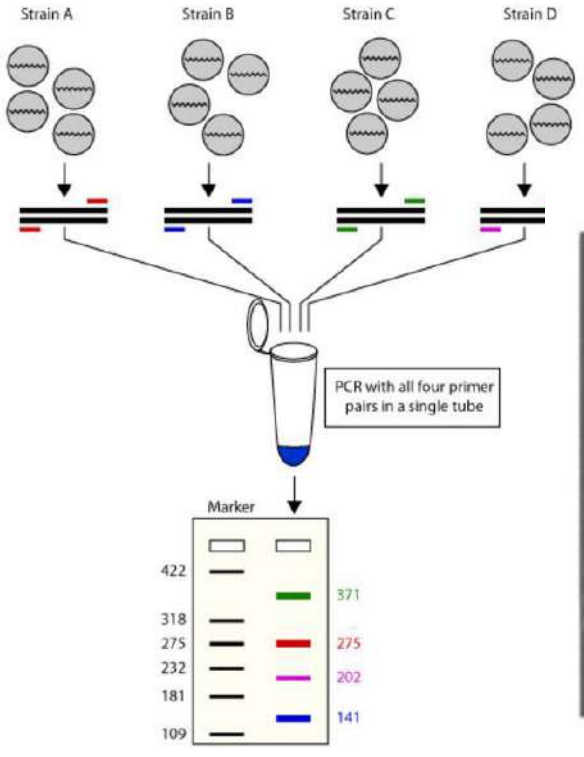


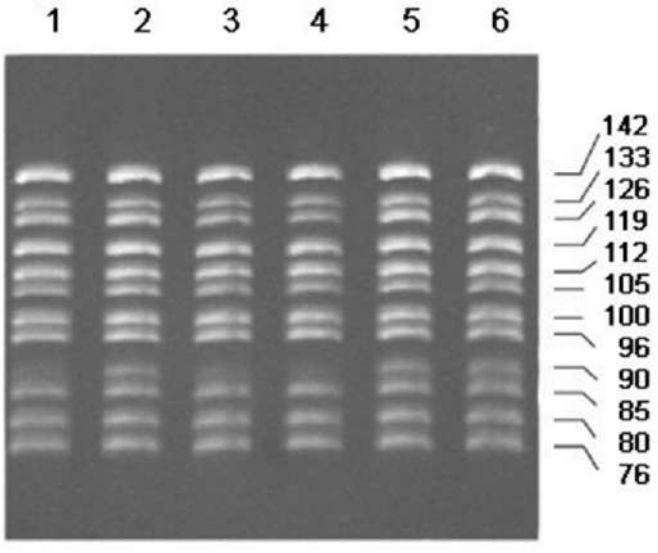
Figura 3.11 Rappresentazione schematica di una reazione ARMS-PCR per lanalisi di una mutazione nota. Per ogni mutazione studiata vengono sintetizzati tre primer, un primer reverse comune e due primer forward allele-specifici che differiscono per il nucleotide allestremità 3 del filamento (sito di mutazione) (a). Lamplificazione nella reazione di PCR dipende dal corretto appaiamento di questa estremità: infatti un mismatch allestremità 3 del primer non permetterà la sintesi del prodotto di amplificazione (b). Per ogni campione vengono allestite due reazioni di PCR utilizzando in ciascuna di queste il primer comune in combinazione con quello corrispondente rispettivamente alla sequenza normale (PCR1) e a quella mutata (PCR2). Lanalisi dei prodotti di amplificazione mediante gel elettroforesi permetterà di discriminare i campioni omozigoti per lallele normale (esclusiva amplificazione in corrispondenza della reazione 1 di PCR), da quelli eterozigoti (amplificazione in corrispondenza della reazione 1 e 2 di PCR) e da quelli omozigoti per la mutazione analizzata (esclusiva amplificazione in corrispondenza della reazione 2 di PCR) (c).



PCR multiplex

E' una tipologia di PCR che utilizza lo stesso DNA per ottenere diversi ampliconi (cioè amplificare contemporaneamente diverse regioni dello stesso genoma) nella stessa reazione. Le PCR Multiplex vanno messe a punto e dosati i primers per ottenere quantità omogenee di amplificati osservabili su gel. Si devono quindi identificare ampliconi che poi possano essere distinti per peso o per qualche caratteristica (come la fluorescenza). E' fondamentale progettare coppie di primers con temperature di melting compatibili e che non diano amplificazioni non corrette (primers specifici). Da ogni reazione si amplifica contemporaneamente più di un amplicone fino a un massimo di quattro o cinque. I criteri principali per questa tecnica sono la compatibilità dei primers, T_m omogenea, assenza di annealing spuri combinati tra coppie diverse, assenza di annealing tra primers, controllo delle amplificazioni desiderate



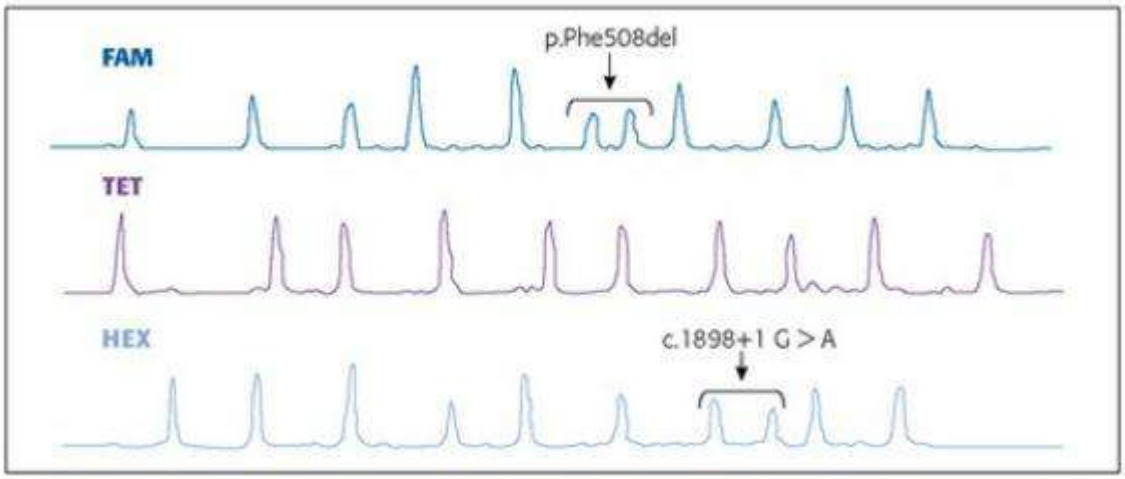


Oligonucleotide Ligation Assay (OLA) o "Saggio di ligazione degli oligonucleotidi"
E' un metodo che utilizza sonde fluorescenti specifiche per le mutazioni analizzate e una reazione di ligazione. E' importante che la sequenza di oligonucleotidi per l'allele mutante sia più lunga di quella del w-t.

Per l'analisi di ogni mutazione si utilizzano tre diverse sonde oligonucleotidiche; due sono complementari rispettivamente all'allele mutato e a quello wild-type e sono associate in posizione 5' a modificatori della mobilità (code di ossido di pentaetilene) di diversa lunghezza; la terza sonda utilizzata è comune, ed è coniugata con un fluorocromo che permette la rivelazione dei prodotti ibridati dopo l'elettroforesi particolare
La DNA ligasi è un enzima termostabile che lega due sequenze adiacenti di DNA se sono perfettamente appaiate a una sequenza di DNA complementare. Per identificare alterazioni puntiformi (per esempio, sostituzioni e piccole delezioni), un segmento di DNA viene amplificato in PCR (PCR multiplex) e successivamente ibridato con sonde complementari alla sequenza amplificata. Si costruiscono oligonucleotidi che si legano a due sequenze adiacenti nel DNA bersaglio, con il punto di giunzione a livello della variante di sequenza di interesse: dopo la reazione di ibridazione tra il DNA in esame e le sonde, la ligasi unirà i due nucleotidi solo se sono perfettamente ibridati. A ogni locus, se la mutazione non è presente, solo l'oligo comune e quello normale daranno un prodotto di ligazione; se è presente solo la mutazione, solo l'oligo comune e mutante daranno prodotto di ligazione; in campioni eterozigoti sia l'oligo normale sia il mutante si legheranno all'oligo comune in uguale quantità, ma dando prodotti di differente mobilità.

La visualizzazione dei risultati avviene con elettroforesi capillare su sequenziatore automatico dotato di un programma che riporta il grafico della corsa elettroforetica e di un altro programma che interpreta i risultati e fornisce il genotipo del campione. Dopo la ligazione, infatti, il prodotto di ogni amplicone ha un'unica combinazione di mobilità elettroforetica e di fluorescenza che permette l'identificazione del genotipo dalla separazione dei picchi dei prodotti di ligazione mediante elettroforesi capillare. Questo si verifica perché le dimensioni degli oligonucleotidi sono tali che i prodotti di ciascuna variante e il corrispettivo allele w-t possono essere differenziati per dimensioni e/o colore del marcatore.

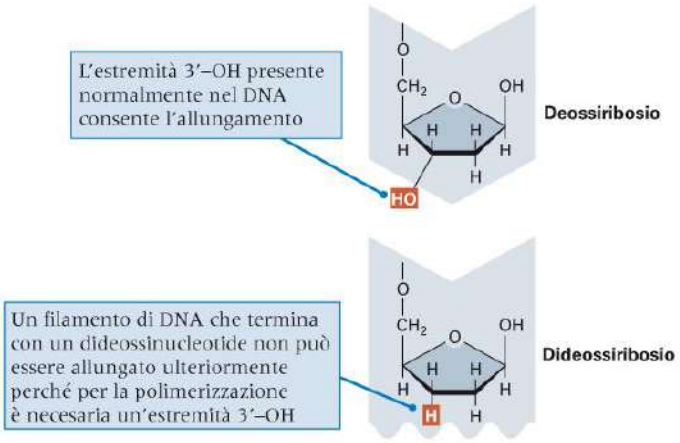
Questa tecnica permette di analizzare contemporaneamente diverse varianti alleliche e rappresenta, per esempio, la tecnica di primo livello nella ricerca di varianti patogenetiche nel gene CFTR, responsabile della fibrosi cistica.



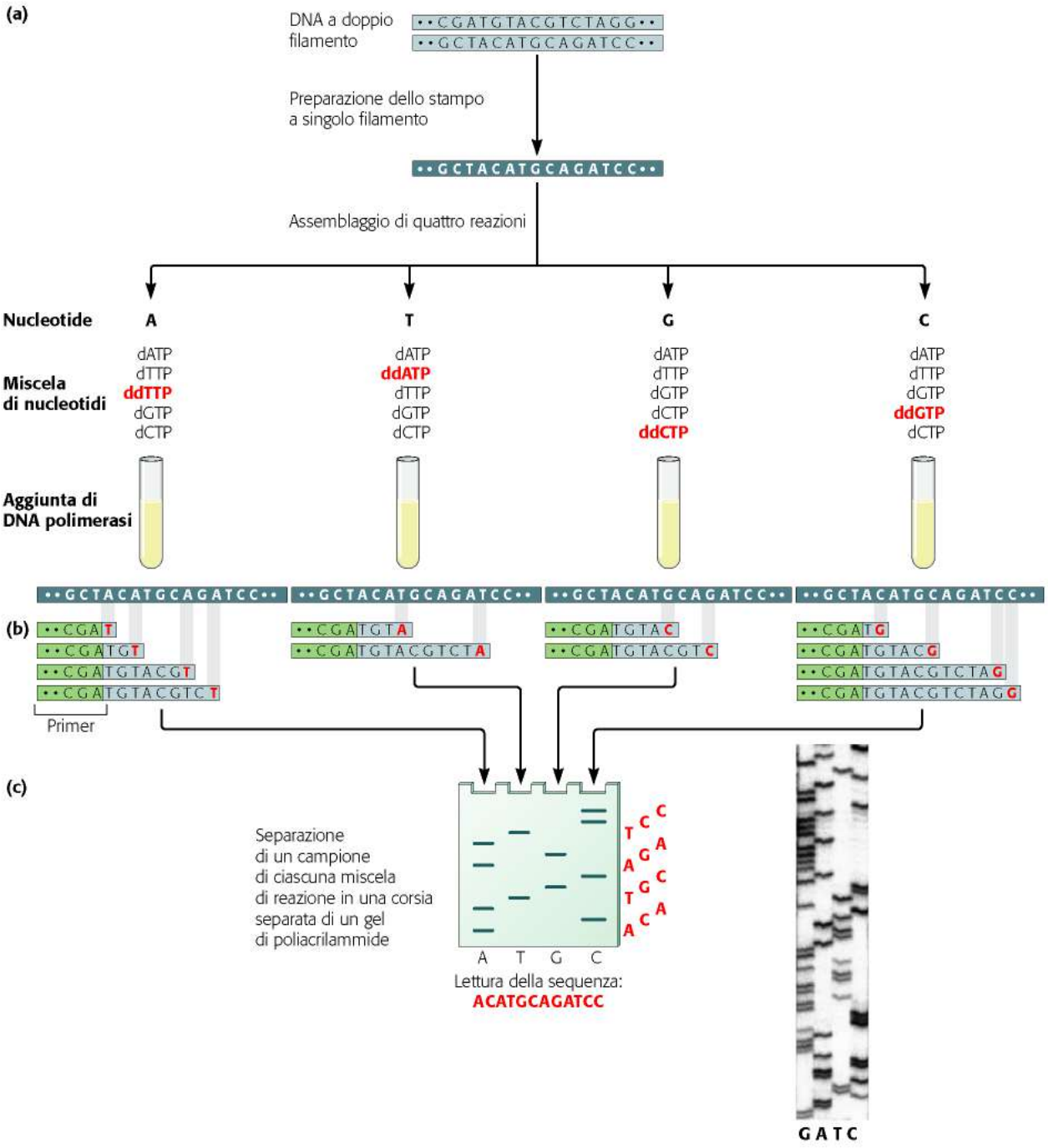
Questo è un elettroferogramma ottenuto dopo una reazione di OLA-PCR in un paziente affettto da fibrosi cistica. Con questo sistema vengono analizzate le 31 mutazioni più frequenti del gene CFTR. Le sonde comuni sono marcate con tre fluorocromi diversi, FAM (blu), TET (rosso), HEX (grigio). L'analisi del DNA del paziente evidenzia la presenza di un doppio segnale corrispondente a due picchi di diverse dimensioni in corrispondenza della mutazione p.Phe508del e della mutazione intronica c.1898+1G>A, dimostrando una condizione di eterozigosi composta per queste due mutazioni.

Il metodo di Sanger (sequenziamento diretto)

Sequenziare il DNA significa determinare l'ordine in cui sono disposte le sue basi. Il metodo dei dideossiterminatori o metodo di Sanger si basa sulla sintesi enzimatica di una nuova catena di DNA su un'elica stampo utilizzando, oltre ai normali deossinucleotidi, dei dideossinucleotidi. Si parte, come per altre tecniche, dall'amplificazione, mediante PCR, del frammento di interesse. Si procede, quindi, alla verifica dell'avvenuta reazione, controllando con l'elettroforesi su gel di agarosio. Si purifica il prodotto di PCR ottenuto (bisogna eliminare i residui di dNTPs e dei primers). Si separano i due filamenti e si inizia la reazione di sequenziamento. Nei primi anni di applicazione di questa tecnica (anni '90), si potevano sequenziare da 100 a 400 bp, mentre dagli anni 2000 in poi, il range si è esteso da 500 a 900 bp. Il sequenziamento si basa su un processo analogo a quello della PCR: si dispone di un primer, che si lega al singolo filamento, avviene la fase di allungamento, che si conclude con quella di terminazione. Analoghi alle quattro basi del DNA, i dideossinucleotidi mancano del gruppo ossidrilico (OH) nel carbonio in posizione 3' dello scheletro di zucchero legato alla base, necessario per formare il legame con il nucleotide successivo e far procedere la neosintesi della catena di DNA; quindi l'incorporazione di un dideossinucleotide provoca l'arresto della reazione polimerasica.



Nel metodo originale i frammenti tronchi prodotti erano radioattivi (o per la presenza nella miscela di reazione di un nucleoside trifosfato radioattivo o perché l'innesco utilizzato era radioattivo) e, dopo elettroforesi, messi in evidenza mediante autoradiografia.

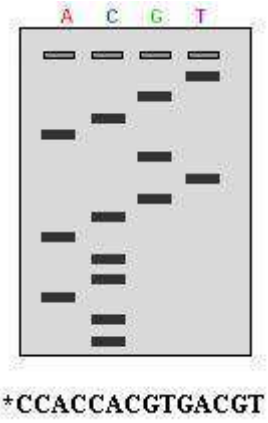


Per il sequenziamento con il metodo di Sanger, nella sua prima versione, si preparano in quattro provette le mix, costituite da:

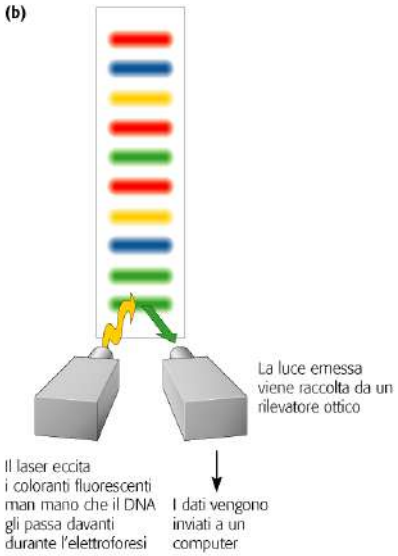
1. Il campione di DNA che si vuole sequenziare a singolo filamento
2. la DNA polimerasi
3. un primer marcato con estremità 3'OH libera
4. i quattro tipi di deossinucleotidi trifosfato normali (dNTP) marcati con 32P o 35S
5. Infine in ogni provetta viene aggiunto uno specifico tipo di dideossinucleotidi trifosfato (ddNTP), in una provetta ddA, in un'altra ddG, in un'altra ddC ed in un'altra ddT, i ddNTP sono la chiave del successo di questa tecnica.

Si verifica che il primer complementa con il filamento stampo e viene esteso dalla DNA polimerasi, questa infatti aggiunge all'estremità libera del primer i nucleotidi complementari al filamento stampo, cioè i dNTP, invece di aggiungere un dNTP di tanto in tanto può accadere che aggiunge un ddNTP. Con l'incorporazione del ddNTP la sintesi del filamento complementare si ferma poiché i ddNTP non permettono la formazione del legame con il successivo nucleotide trifosfato. I dNTP hanno un 3'OH nello zucchero desossiribosio che consente l'aggiunta di nucleotidi, i ddNTP invece hanno un 3'-H che non consente l'aggiunta di ulteriori nucleotidi e per questo la sintesi del DNA termina.

La stessa cosa accade per tutti i frammenti di ciascuna provetta ovvero si ha la sintesi del filamento complementare fino a che non viene incorporato un ddNTP. Si ottengono in questo modo frammenti di DNA di varia lunghezza, che differiscono da provetta a provetta per il tipo di ddNTP con il quale terminano. I campioni marcati contenuti nelle quattro mix vengono poi denaturati al calore e caricati in quattro pozzetti diversi, in un gel di poliacrilammide. L'elettroforesi viene realizzata a circa 70°C in presenza di urea in modo da impedire la rinaturazione del DNA. Con l'elettroforesi i frammenti migrano in base alla dimensione e quindi al termine della elettroforesi, mediante autoradiografia, è possibile risalire alla sequenza.

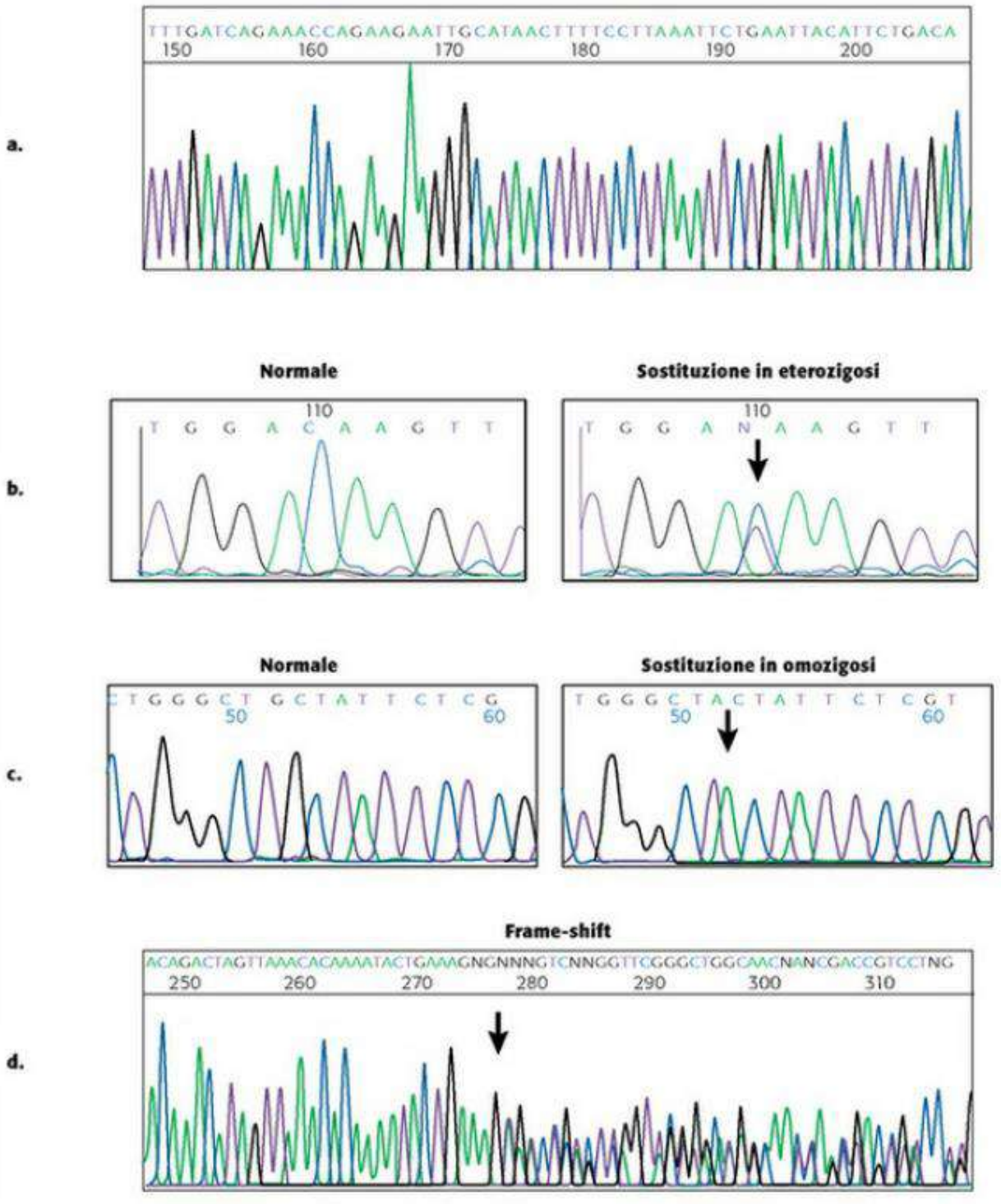


La sequenza viene letta dal basso del gel verso l'alto. Nella prima corsia elettroforetica (A), si osservano 3 bande, questo significa che in quelle posizioni è stato incorporato un ddATP e quindi significa che in quelle 3 posizioni c'è una A nella sequenza. Si procede con lo stesso ragionamento per le altre 3 corsie elettroforetiche che sono indici di dove sono posizionati i nucleotidi C, T, G. In basso è riportata la sequenza del filamento.



Attualmente (dagli anni 2000 in poi) esistono metodi di sequenziamento automatico, la sintesi dei filamenti procede come già descritto prima, però con questa metodica vengono incorporati, per terminare la sequenza, ddNTP marcati con fluorocromi diversi. I campioni così ottenuti vengono separati in funzione della loro lunghezza tramite elettroforesi capillare sul sequenziatore automatico e, quando raggiungono la posizione del detector, vengono rilevati e identificati grazie all'emissione a diverse lunghezze d'onda dei fluorocromi utilizzati come

marcatori. Le emissioni vengono raccolte e analizzate da una fotocamera digitale CCD che riporta la loro sequenza in un grafico chiamato elettroferogramma, caratterizzato da una successione di picchi di quattro colori (adenina in verde, citosina in blu, guanina in nero e timina in rosso), corrispondenti ai diversi fluorocromi. La procedura è completamente automatizzata ed in questo modo si ha un grosso risparmio in termini di tempo, poiché si possono analizzare più campioni simultaneamente.



Sequenziamento Sanger. (a) Esempio di un elettroferogramma ottenuto per sequenziamento automatico. Ogni picco, con uno pseudocolore diverso (rosso-timina, verde-adenina, nero-guanina, blu-citosina), corrisponde a una base della sequenza nucleotidica.

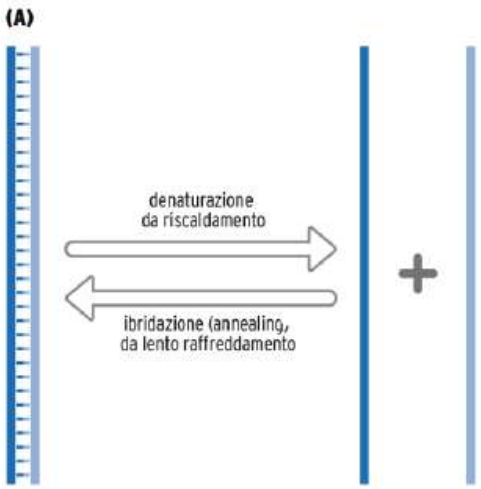
(b) Presenza di una sostituzione di base (C>T) in eterozigosi: si evidenzia per la contemporanea presenza di due picchi nella stessa posizione. (c) Presenza di una sostituzione di base (G>A) in omozigosi: si evidenzia come la presenza di un picco singolo diverso da quello presente nella sequenza normale, corrisponde al nucleotide sostituito. (d) Presenza di una mutazione frameshitf in eterozigosi. Dal punto indicato dalla freccia si sovrappongono due sequenze, corrispondenti all'allele normale e a quello mutato. A = adenina; C = citosina; G = guanina; N = presenza contemporanea di più basi in una stessa posizione, indicativa di una condizione di eterozigosi; T = timina.

I principi di ibridazione degli acidi nucleici

In una molecola di DNA a doppio filamento, i legami a idrogeno fra le basi appaiate si comportano come una "colla" che tiene insieme i due filamenti di DNA complementare. Due legami a idrogeno si formano tra A e T, tre invece sono quelli che uniscono le G alle C. Una regione ricca di G e C è, perciò, più stabile di una ricca di A e T.

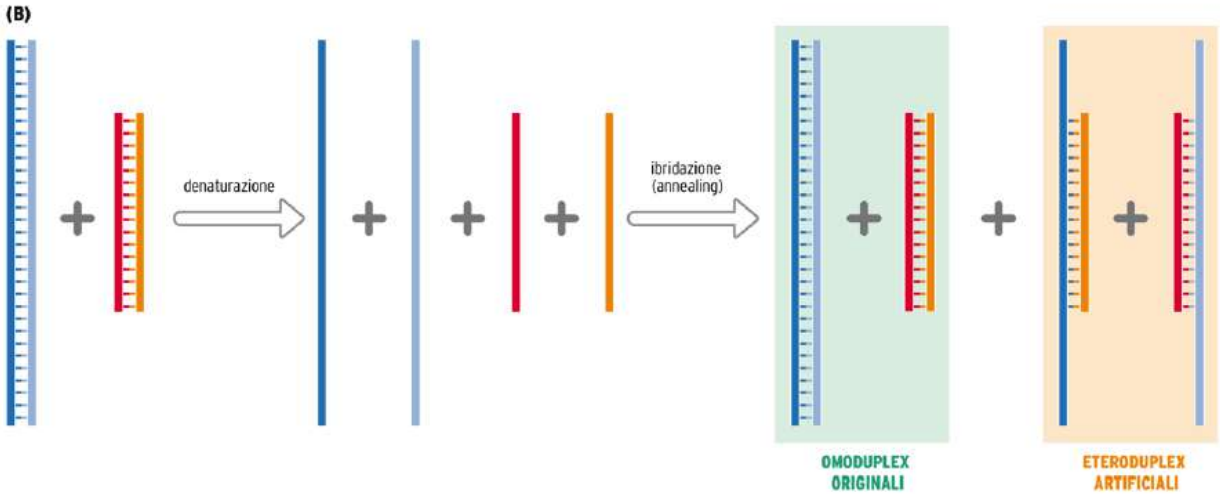
Ogni legame a idrogeno, preso da solo, è debole, ma, quando l'appaiamento delle basi si estende per un tratto molto ampio, la forza risultante dei legami a idrogeno diventa piuttosto intensa (un po' come un singolo filo di velcro che da solo non fa presa, ma la somma di tanti fili permette di legare e aderire a una superficie garantendo un'ottima resistenza).

Il DNA a doppio filamento può essere sottoposto a diversi trattamenti che provocano la rottura dei legami a idrogeno, causando la separazione dei due filamenti di DNA (denaturazione del DNA). Per esempio, se si porta il DNA ad alte temperature o lo si espone ad un'elevata concentrazione di una molecola estremamente polare, come la formammide o l'urea, i legami a idrogeno si rompono e i due filamenti complementari di DNA si separano. Il graduale raffreddamento che segue la denaturazione permette ai filamenti di DNA separati di riassociarsi, formando nuovamente gli appaiamenti di basi nell'ordine corretto per ripristinare il doppio filamento originale.



L'associazione di due qualsiasi filamenti complementari di acidi nucleici a formare un acido nucleico a doppio filamento prende il nome di ibridazione (o annealing) degli acidi nucleici. In condizioni sperimentali, due filamenti singoli di acidi nucleici con un alto livello di complementarità delle basi possono essere in grado di ibridarsi a formare un duplex artificiale. Per esempio, se si mescolano frammenti di DNA a doppio filamento clonati che provengono da due "sorgenti" diverse, ma presentano un'elevata percentuale di identità, e si scalda il composto fino a rompere i legami a idrogeno, tutti i milioni di molecole di DNA a doppio filamento, nel DNA di ciascuna delle due "sorgenti", verranno resi a singolo filamento.

Ora si immagini di poter lasciar raffreddare il composto lentamente: si formano due diversi tipi di DNA duplex. Nel primo caso, una parte delle molecole di DNA a singolo filamento si appaierà con il filamento partner originale a ricostituire i filamenti di DNA originali (omoduplex). In aggiunta, però, talvolta, una molecola di DNA a singolo filamento si appaierà a un filamento di DNA complementare proveniente dall'altra "sorgente" a formare un eteroduplex artificiale. Il termine eteroduplex viene usato per indicare tutti i duplex artificiali nei quali l'appaiamento delle basi non è perfetto per tutta la lunghezza dei due filamenti complementari. Nell'esempio riportato in figura si osserva un matching perfetto lungo tutti e due i brevi filamenti di DNA, ma gran parte dei filamenti in blu rimane non appaiata. Molto raramente filamenti di DNA complementari provenienti da due diverse "sorgenti" possono essere progettati in modo da avere identica lunghezza e perfetto appaiamento di basi – altrimenti, potrebbero formare omoduplex artificiali.



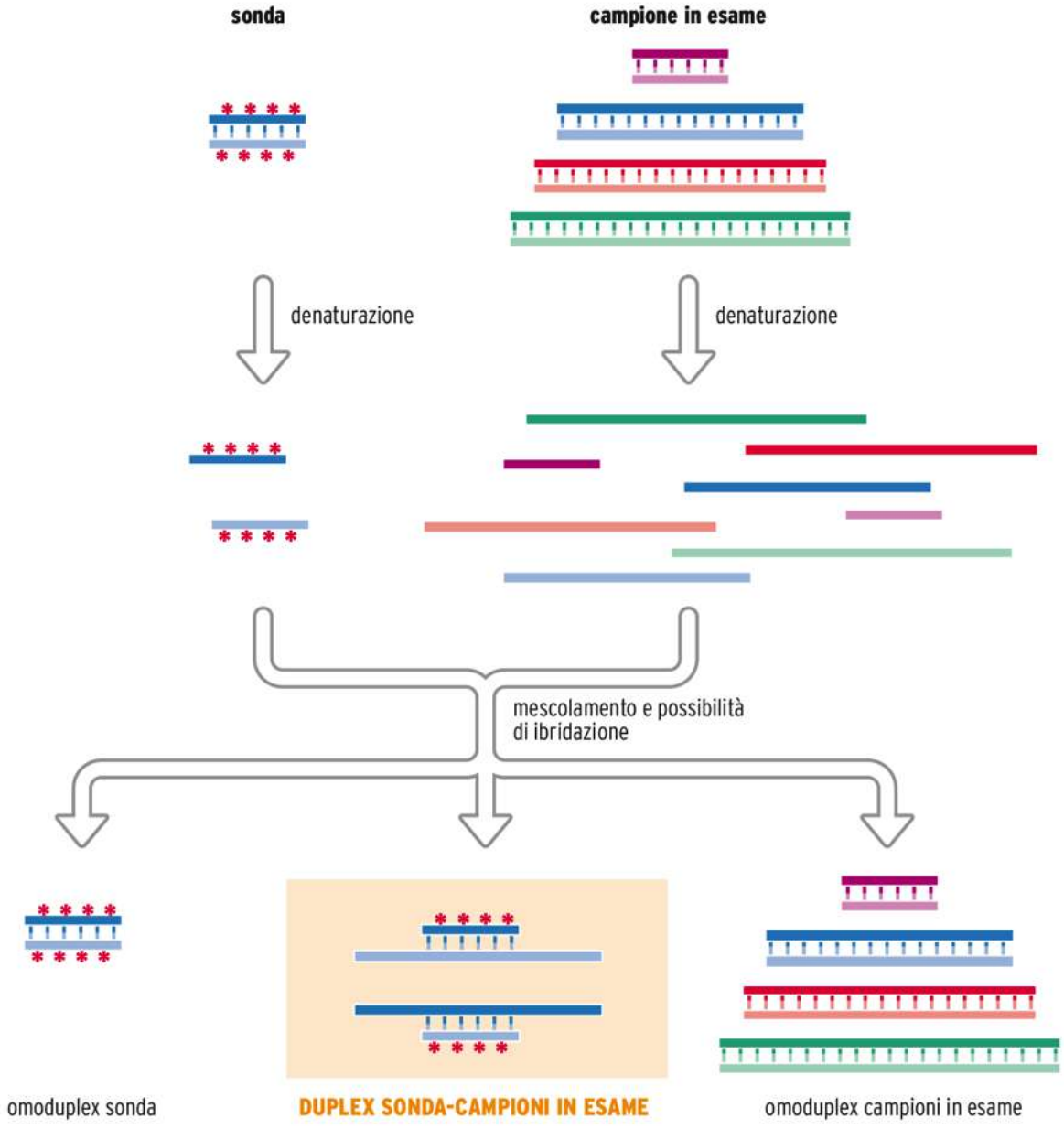
La formazione di duplex artificiali, quasi sempre eteroduplex, è l'essenza dei saggi di ibridazione nucleica che sono ampliamente usati nella genetica molecolare. Per comodità sono stati illustrati DNA a doppio filamento clonati, ma spesso accade che la molecola di partenza tra gli acidi nucleici sia l'RNA (solitamente già a singolo filamento) oppure siano oligonucleotidi sintetici, ma anche il DNA stesso. Spesso, inoltre, uno o entrambi gli acidi nucleici sono mix composti di frammenti, come l'RNA di una cellula o frammenti dell'intero DNA genomico. Come il DNA clonato, gli acidi nucleici di partenza sono, di norma, isolati da milioni di cellule, quindi le singole sequenze sono presenti in un numero cospicuo di copie, spesso in milioni di copie.

L'ibridazione degli acidi nucleici sfrutta la specificità dell'ibridazione. Due filamenti singoli di polinucleotidi (DNA o RNA) o di oligonucleotidi formeranno un ibrido a doppio filamento, stabile (duplex) solo se c'è una percentuale alta di appaiamento tra di loro. La stabilità dell'ibrido risultante dipende dalla lunghezza dei tratti di perfetta complementarità. Le condizioni del saggio possono essere scelte in modo da permettere soltanto duplex con matching perfetti oppure ammettere una certa quantità di mismatch.

I saggi di ibridazione possono essere realizzati in molti modi diversi, con varie applicazioni sia nella ricerca sia nella diagnostica. C'è, comunque, un principio comune che sta alla base: una popolazione

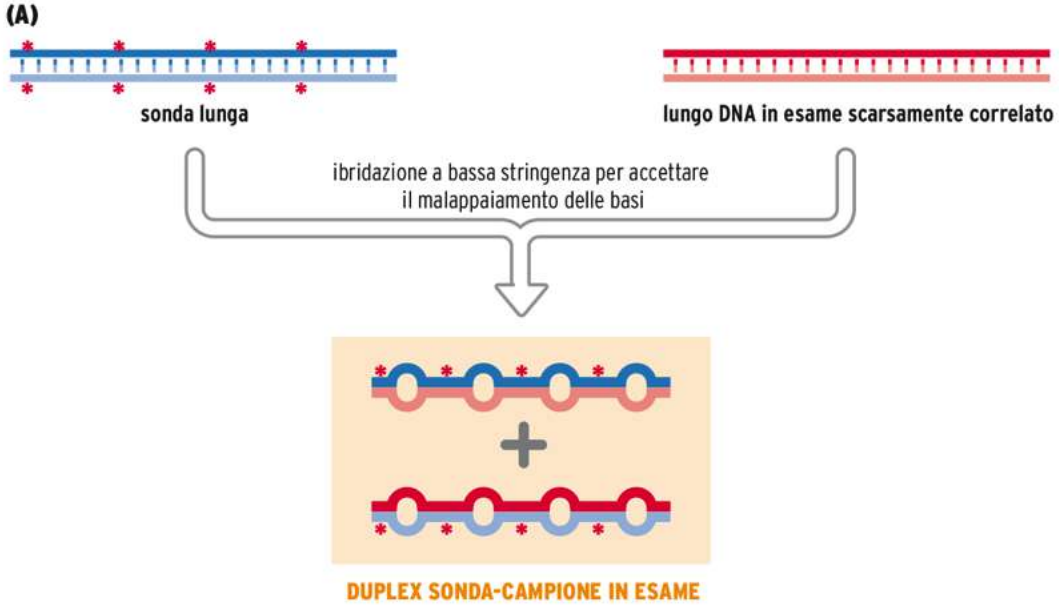
nota, ben distinta di molecole di acidi nucleici o di oligonucleotidi sintetici (si parla di popolazione di sonde) viene utilizzata per studiare una popolazione non ancora completamente compresa di acidi nucleici (il campione in esame). Per arrivare a questo scopo, è necessario separare entrambe le popolazioni (sonde e campioni) in filamenti singoli e poi mischiarli cosicchè i filamenti singoli delle sonde possano formare duplex artificiali con i filamenti a loro complementari provenienti dal campione.

Dal momento che l'oggetto del saggio di ibridazione è l'uso di sonde per identificare filamenti complementari, interamente o parzialmente, al campione, le sonde che si legano al campione devono essere etichettate in modo da poter essere rintracciate. Per fare ciò, o la sonda o il campione necessitano di essere etichettati fin dall'inizio.

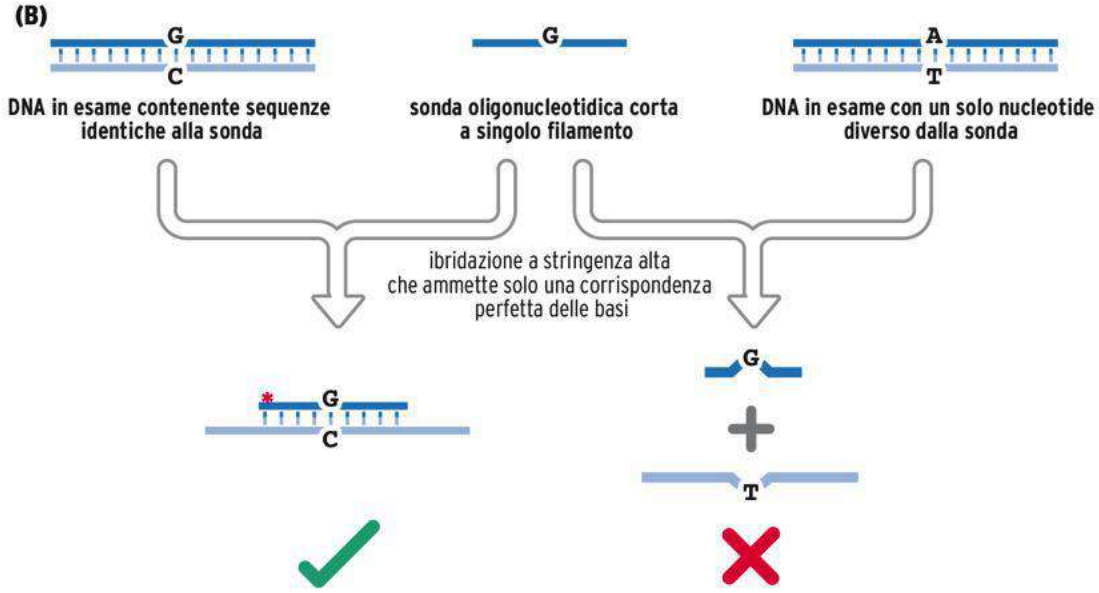


Il saggio di ibridazione può essere usato per identificare sequenze di acidi nucleici che sono "imparentati" a distanza di una data sonda di acidi nucleici. Si potrebbe partire da un clone di DNA da un gene umano e usarlo per identificare il corrispondente gene, ma nel topo.

Il gene umano potrebbe essere significativamente diverso da quello murino dal punto di vista della sequenza, ma se si sceglie una sonda di DNA lunga e riduciamo la stringenza di ibridazione, sarà possibile la formazione di eteroduplex stabili, anche qualora si presentassero numerosi mismatch.



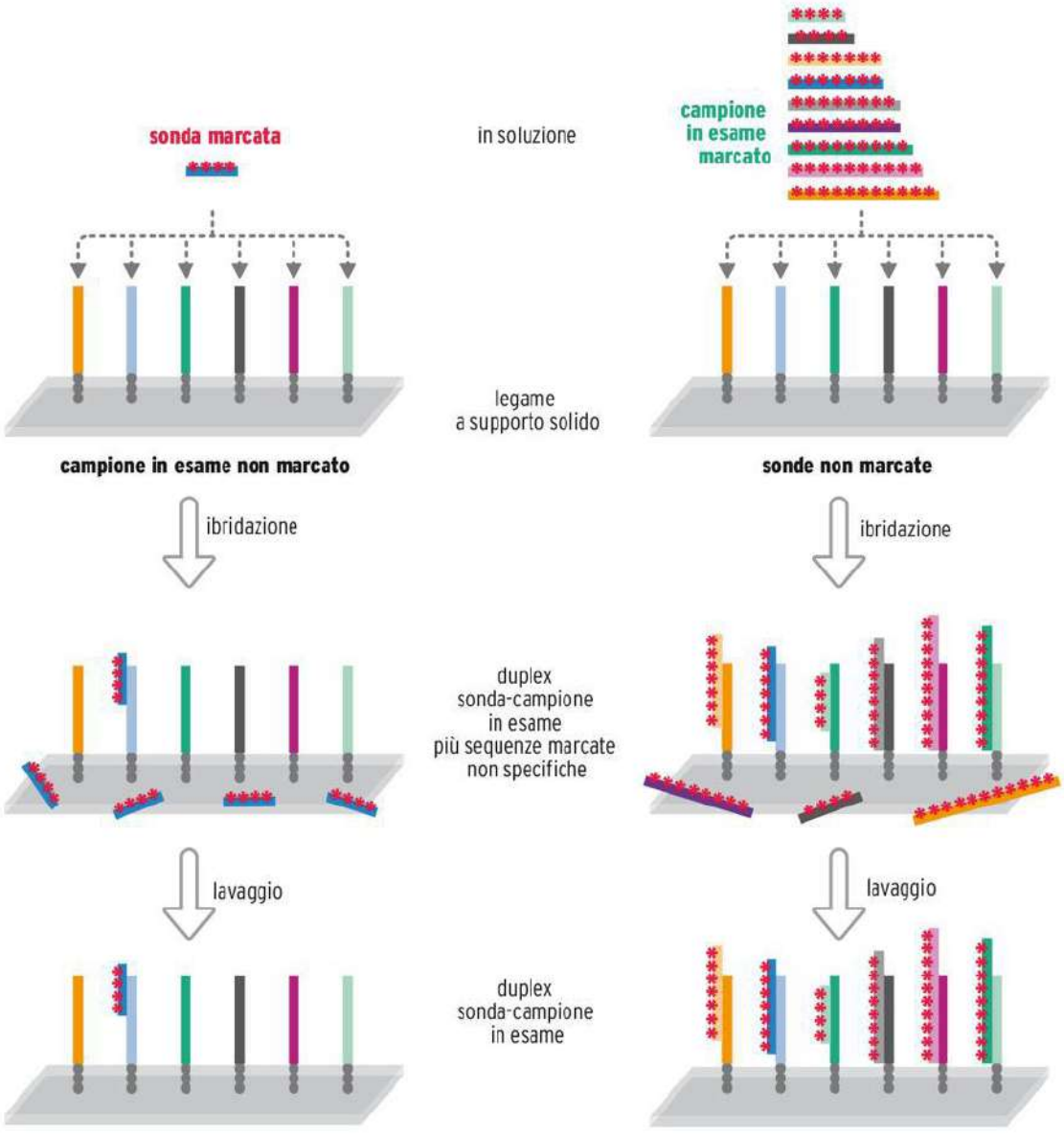
Analogamente, si può optare per condizioni di ibridazione che ammettano soltanto perfetti matching. Se si sceglie una sonda oligonucleotidica, si predispone un'alta stringenza di ibridazione in modo che i duplex sonda-campione che si possono formare sono quelli che contengono esattamente la stessa sequenza della sonda. Questo può verificarsi perché un singolo mismatch su una sequenza totale di, per esempio, 18 nucleotidi, può rendere il duplex instabile termodinamicamente. Gli oligonucleotidi, perciò, possono essere usati per identificare gli alleli che si differenziano per un singolo nucleotide (si parla di oligonucleotidi allele-specifici).



Esistono diversi tipi di saggio di ibridazione, ma tutti quanti rientrano in due ampie classi.

In un caso, le molecole sonda sono etichettate e le molecole del campione non lo sono, come nella figura sopra. In quel caso, la sonda è spesso un tipo a singolo filamento di DNA clonato e viene solitamente etichettata utilizzando una polimerasi che sintetizzi il filamento complementare di RNA o DNA in presenza di uno o più nucleotidi etichettati con fluorocromi. L'altra tecnica si serve di sonde non etichettate, mentre le molecole del campione sono etichettate.

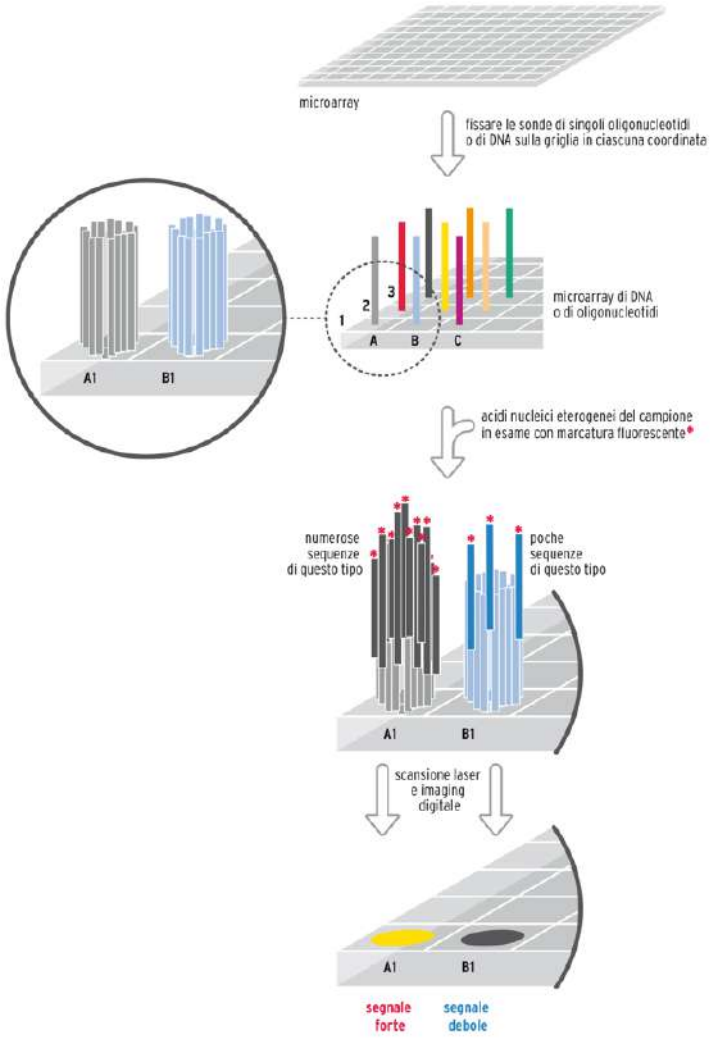
Il motivo per cui si usano acidi nucleici etichettati in un saggio di ibridazione è riuscire a identificare molecole eteroduplex sonda-campione. Come si può distinguere l'etichetta di questi duplex da quella originale della sonda o da quella del DNA del campione in esame? La risposta consiste nell'immobilizzare la popolazione di acidi nucleici non etichettata su un supporto solido (spesso di plastica, vetro, o quarzo), ed esporla ad una soluzione acquosa contenente una popolazione di acidi nucleici etichettati. Quando i filamenti di acidi nucleici etichettati si ibridano con le sequenze complementari sul supporto solido, essi si legheranno fisicamente al supporto, ma le molecole etichettate che non trovano un partner sul supporto o che si attaccano in maniera non specifica, possono essere lavate via. Ciò lascia al suo posto i partner complementari per i quali il saggio è stato progettato.



I saggi di ibridazione sono usati per vari scopi. Per decine di anni, quasi tutti i saggi di ibridazione hanno usato sonde (spesso, cloni di DNA a singolo filamento) etichettate in maniera omogenea, alla ricerca di sequenze collegate, in un campione complesso di acidi nucleici immobilizzato su un supporto. Come descritto di seguito, i saggi di ibridazione su microarray sono diventati molto popolari; si servono di popolazioni di sonde complesse non etichettate, legate a una superficie per indagare un campione etichettato (vedi figura precedente, colonna di destra).

L'ibridazione su microarray

Lo sviluppo di potenti e innovative tecnologie di ibridazione nei primi anni '90 permette la conduzione simultanea di numerosi saggi di ibridazione su uno stesso campione alle stesse condizioni. Un microarray ("micro schieramento") di DNA o di oligonucleotidi è composto da diverse migliaia o milioni di diverse popolazioni di sonde di DNA o di oligonucleotidi non etichettate, che sono state fissate su una superficie di vetro o di un altro materiale adatto, all'interno di una griglia disegnata ad alta densità. All'interno di ciascun quadretto della griglia ci sono milioni di copie identiche di una sola sonda (un quadretto della griglia con la propria popolazione di sonde viene detto "feature"). Per esempio, i microarray oligonucleotidici hanno spesso una superficie di 1,28 cm x 1,28 cm che contiene milioni di diverse features, ciascuna di dimensioni di circa 5 o 10 μm^2.



Una quantità di campione – una soluzione acquosa contenente una popolazione complessa di DNA o RNA denaturati etichettati con fluorocromi – viene ibridata con le diverse popolazioni di sonde sul microarray. Un array di DNA può essere "spottato" da un array di "penne" in una singola operazione di scrittura. Dopo un passaggio di lavaggio per rimuovere le molecole che si sono legate non specificamente al campione etichettato sull'array, la rimanente etichetta legata fluorescente viene passata attraverso uno scanner laser ad alta risoluzione. Il segnale emesso da ogni feature presente sull'array viene analizzato da un software di imaging digitale che converte il segnale fluorescente di ibridazione in uno scelto tra una palette di colori a seconda dell'intensità del segnale ricevuto.

Dal momento che l'intensità di ciascun segnale di ibridazione riflette il numero di

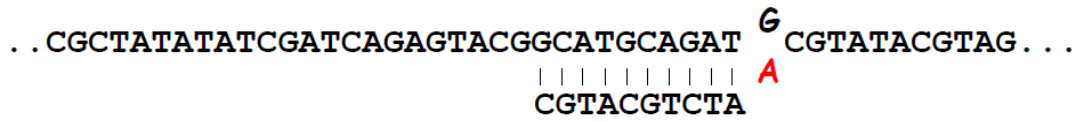
molecole etichettate che si sono legate a una feature, l'ibridazione su microarray viene utilizzata per quantificare diverse sequenze in popolazioni complesse del campione, come diversi estratti di DNA genomico o di RNA totale di una cellula (o di cDNA). Applicazioni frequenti comprendono la quantificazione di trascritti diversi (per trovare i profili di espressione genica) e anche lo scanning di genomi alla ricerca di delezioni o duplicazioni su larga scala.

APEX (Arrayed Primer Extension)

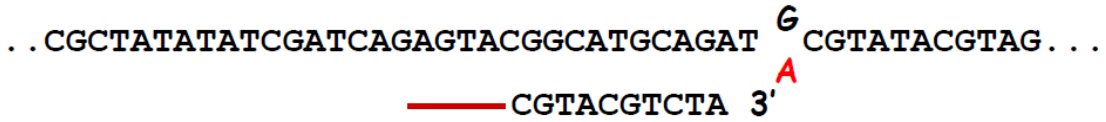
Legata alla messa a punto dell'ibridazione su microarray è la tecnica denominata APEX, il cui scopo è di estendere un primer fluorescente allele-specifico, che si leghi perfettamente al campione inserito nel supporto e che possa essere facilmente letto da un computer, al fine di elaborarne l'informazione.

Si procede così:

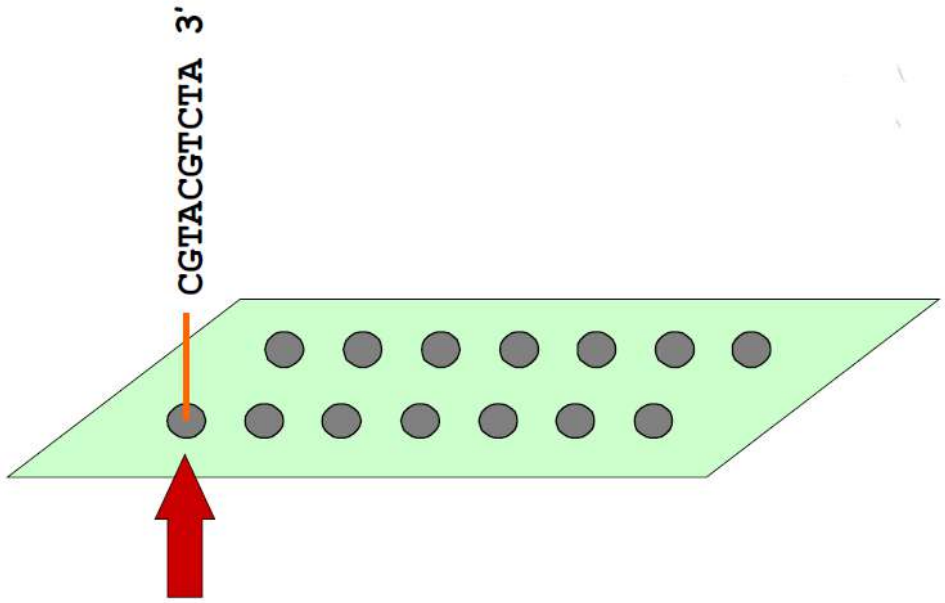
1. Si disegnano gli oligonucleotidi complementari alla sequenza fiancheggiante la mutazione/polimorfismo;



2. Si sintetizzano gli oligonucleotidi con un 5' linker;

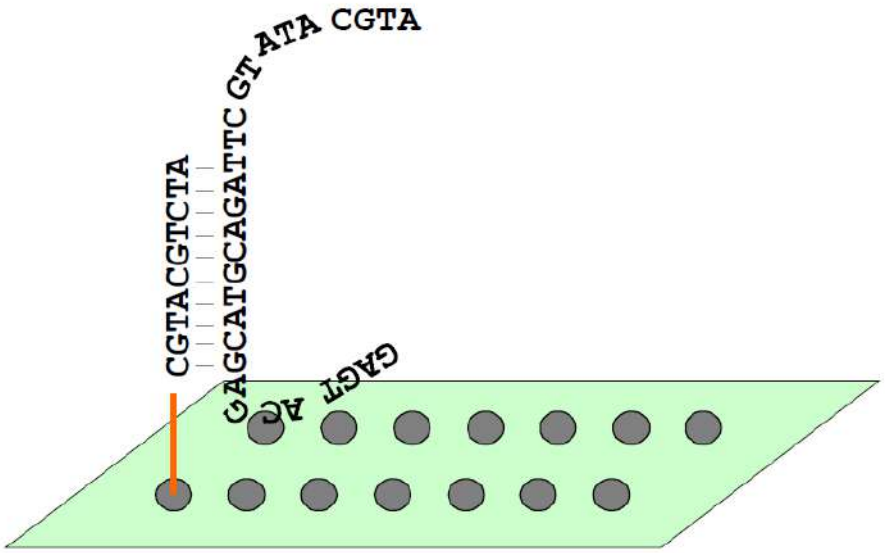


3. Si dispongono gli oligo in maniera sito-specifica. La superficie del supporto di solito viene trattata con amminosilano che favorisce l'adesione tra supporto e acido nucleico;

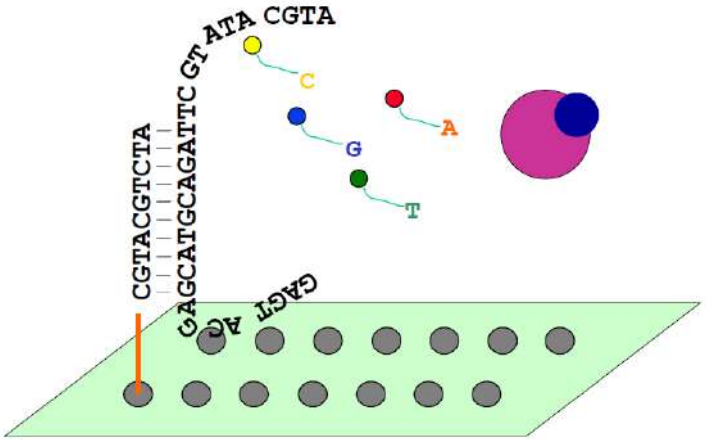


4. Si fa avvenire l'ibridazione con i prodotti di PCR;

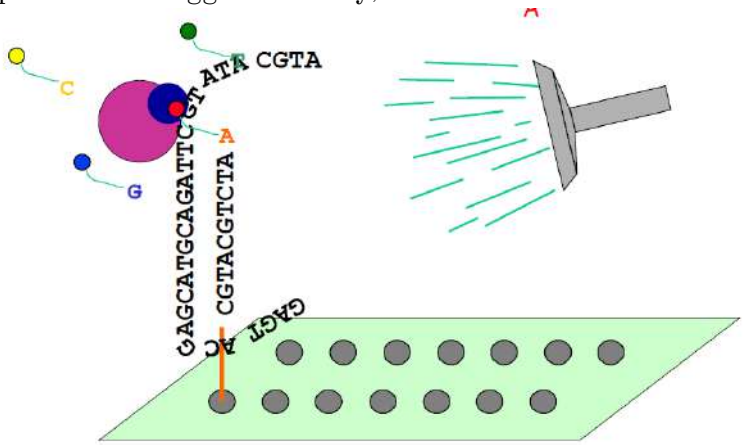
..CGCTATATATCGATCAGAGTACGGCATGCAGAT G/A CGTATACGTAG...



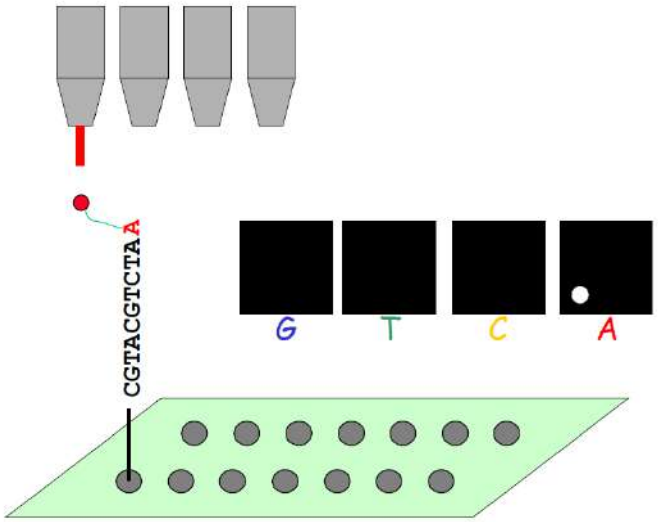
5. Si estende il primer con dideossinucleotidi, legati ad un fluorocromo, di colore diverso per ciascuna base (la pallina è la DNA polimerasi);



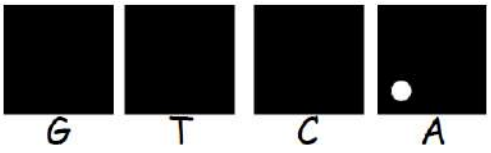
6. Si procede al lavaggio dell'array;



7. Si effettua la lettura del filamento con uno scanner a 4 colori;

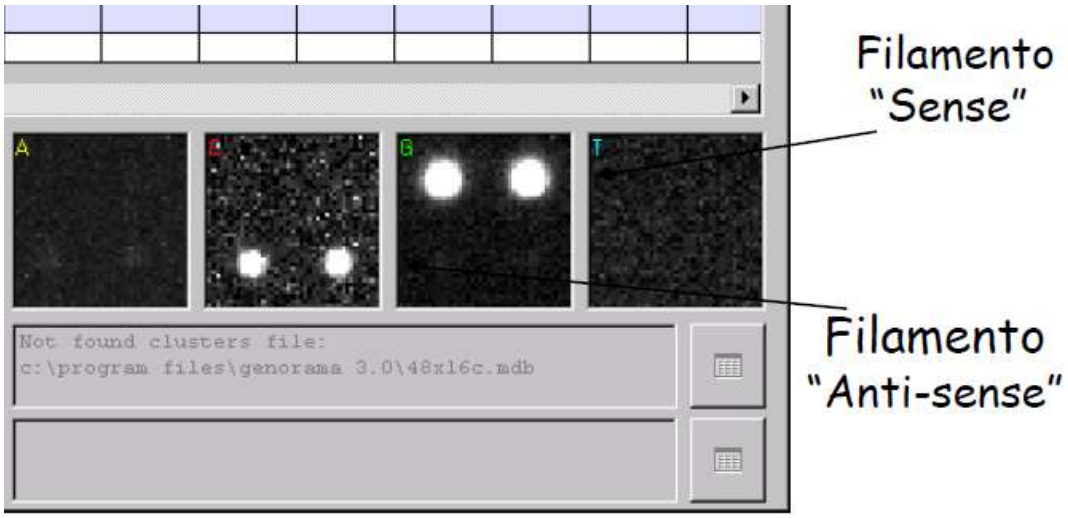


8. Il software procede alla trasformazione automatica dei dati letti in genotipi.



DNA#	Calls G	T	C	A	Possible alleles	Genotype
XYZ	0	0	0	100	G or A	A

Il computer è anche in grado di mostrare quale base è localizzata sul filamento senso e quale sull'antisenso.



La procedura si conclude con la convalida degli array sviluppati prima con templati sintetici poi con campioni di DNA genomico.

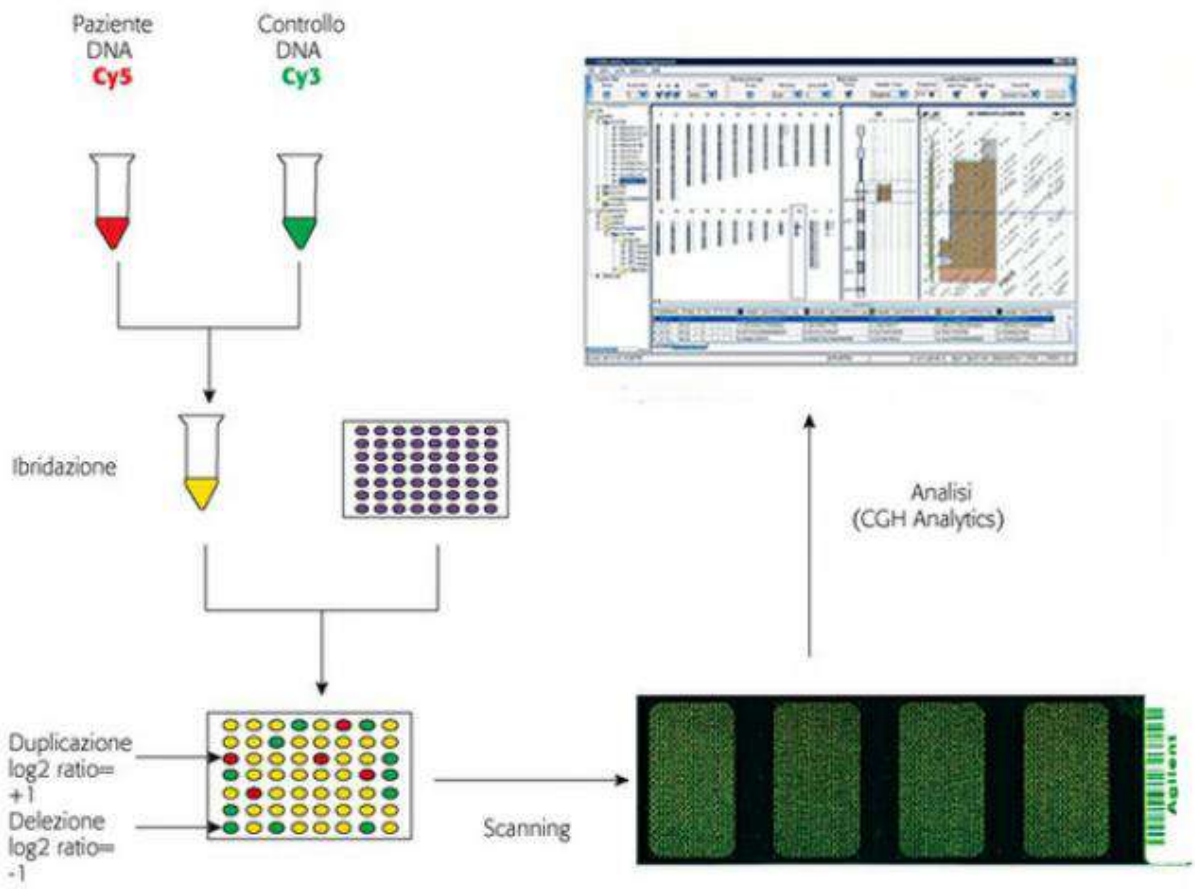
In tutte le tecniche di polimerizzazione viste è possibile ricorrere a primers mutagenici, cioè primer non perfettamente complementari alla sequenza presa in esame. I primer vengono così progettati in modo che l'amplificato, ad esempio, codifichi per un amminoacido diverso dal w-t. Ciò si ottiene "inducendo" una mutazione nei filamenti che verranno trascritti e duplicati, che, ad esclusione di quelli del campione, saranno tutti portatori della "mutazione" desiderata. Il meccanismo attraverso il quale si può ottenere questo è identico a quello dell'amplificazione dei falsi bersagli.

Un esempio di applicazione di APEX è descritto in uno studio pubblicato nel 2006 su "Pediatrics". E' stato provato un approccio diagnostico per lo screening dei neonati nuovo (per quel momento), basato sui microarray, alla ricerca di mutazioni multigeniche simultanee che provocavano la perdita di udito sensorineurale. La perdita dell'udito, in effetti, è un difetto molto comune alla nascita. Nelle nazioni industrializzate, circa un bambino su 1000 nasce sordo, altri individui lo diventano nell'età adulta. In totale, la perdita di udito colpisce dal 6 all'8% della popolazione. La perdita dell'udito può essere la conseguenza di fattori ambientali, eziologia genetica, o entrambi. La base genetica della perdita dell'udito è, però, straordinariamente complessa. La perdita di udito genetica può seguire un pattern autosomico dominante, autosomico o X-linked recessivo o di ereditarietà mitocondriale. Diverse mutazioni a carico dello stesso gene possono causare perdita di udito dominante o recessiva, e in alcuni casi la sordità può essere causata da mutazioni in due geni diversi dello stesso gruppo funzionale. Per compiere le operazioni di diagnosi, è stato sviluppato un saggio su microarray, che attraverso l'ibridazione seguita dall'aggiunta di un singolo nucleotide (cioè un'APEX), può determinare la base genetica molecolare della perdita di udito sensorineurale non sindromica e/o sindromica (SNHL) a causa di mutazioni in svariati geni della connettina, di mutazioni in 2 geni della famiglia SLC26 e mutazioni in 2 geni mitocondriali che contribuiscono alla perdita dell'udito. I primer w-t, per il filamento senso e per quello antisenso, utilizzati vennero disegnati secondo le sequenze depositate in Genbank. Le conclusioni dello studio furono le seguenti: test genetici per la perdita di udito sensorineurale, come quello appena terminato, completo e relativamente non costoso, avrebbe migliorato la gestione medica dei pazienti affetti e il counseling genetico per le loro famiglie. L'APEX, in generale, presenta il considerevole vantaggio di esaminare più loci di uno stesso genoma e più genomi contemporaneamente.

Array-CGH

La tecnica array-CGH ("array-based Comparative Genomic Hybridazation") consiste nell'ibridazione del DNA del soggetto in esame con un DNA di controllo, rappresentativo del genoma della specie, ciascuno marcato con un diverso fluorocromo, su piccole sequenze di DNA disposte su un vetrino, a costituire appunto un "array". Le sequenze possono essere degli oligonucleotidi specifici per gli esoni dei geni trascritti, oltre che per molti introni, e in questo caso si parla di "oligonucleotide-array", e possono variare di numero a seconda della distanza genomica da cui sono intervallate, cioè a seconda della risoluzione che si intende prestabilire per l'array (3,5 – 100 kb). Se il DNA di controllo e il DNA testato sono equimolecolari, essi si annullano in qualche modo a vicenda e non si apprezzeranno

differenze di ibridazione, in eccesso o in difetto, del DNA testato rispetto al DNA di controllo. Al contrario, in presenza di una delezione o di una duplicazione, le elaborazioni successive alla co-ibridazione devieranno in maniera specifica, nel senso che il segnale di ibridazione è a favore del DNA di controllo nel caso di microdelezione, e a favore del DNA testato nel caso di microduplicazione.

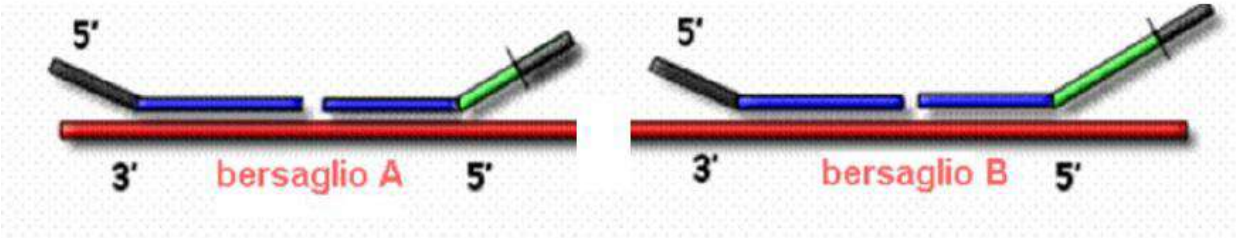


Il DNA del paziente (Cy5) è marcato in rosso, il DNA di controllo è marcato in verde (Cy3). Entrambi i DNA vengono coibridizzati sul vetrino dove sono organizzate le sonde molecolari che costituiscono l'array, e il segnale di ibridazione viene letto da un computer dedicato: se prevale il rosso (+1 in scala logaritmica significa che il paziente è portatore di una duplicazione relativamente al locus testato dalla sonda; se prevale il verde (-1 in scala logaritmica) significa che il paziente è portatore di una delezione. Un'ulteriore elaborazione grafica operata dal computer dedicato consente di visionare simultaneamente la banda cromosomica a cui si è verificato l'arrangiamento, l'estensione del riarrangiamento stesso e il suo contenuto genico.

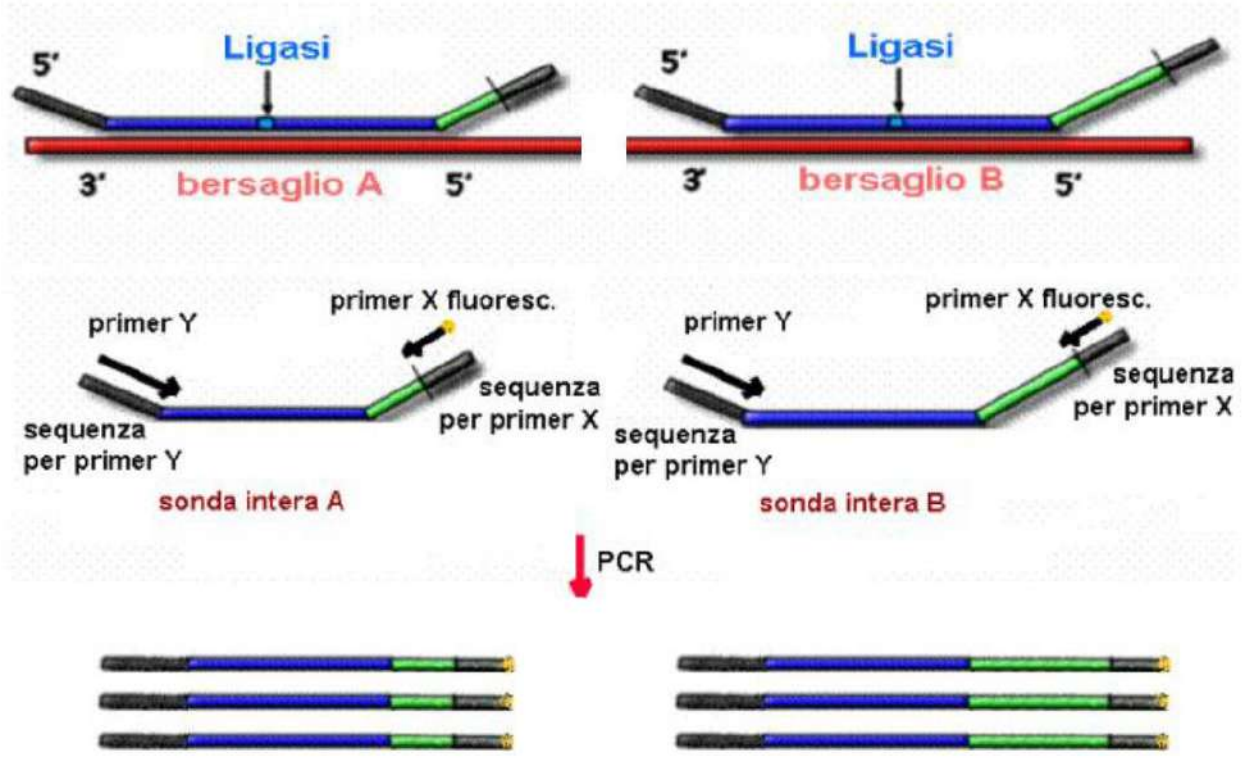
MLPA (Multiplex Ligation Probe Assay)

Il metodo utilizza due sonde oligonucleotidiche per ogni target di interesse.

Le sonde sono costruite in modo da essere contigue nella regione da analizzare e contengono all'estremità 3' e 5' sequenze che serviranno per i primer, X e Y, uguali per tutte le sonde in esame; inoltre, una delle due sonde contiene una sequenza di lunghezza variabile per analisi contemporanea di più target (sequenza "stuffer").

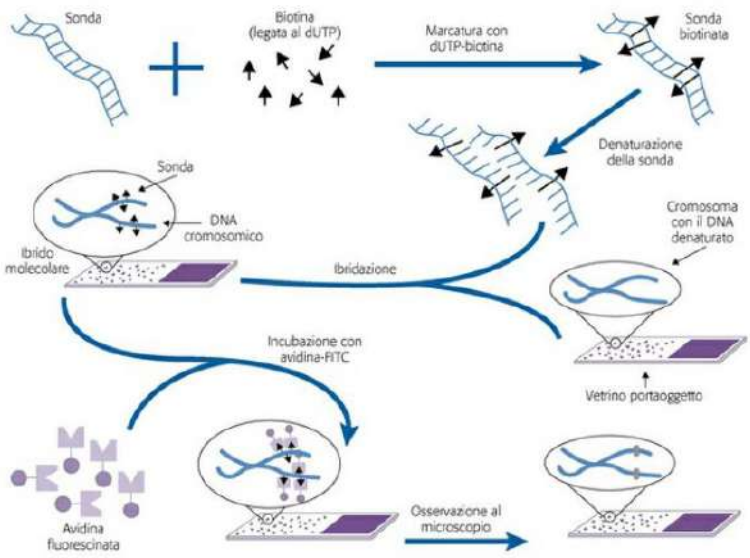


La reazione prevede l'ibridazione delle sonde al DNA target (che avviene in modo stechiometrico, cioè direttamente proporzionale alla quantità della sequenza target), seguita da una reazione di ligazione delle due sonde a opera della DNA ligasi. Una reazione di PCR utilizzando i primer X e Y (uno dei quali è coniugato con un marcatore fluorescente) amplificherà esclusivamente i prodotti ottenuti dalla ligazione, che verranno separati mediante elettroforesi capillare su sequenziatore automatico.



Frammenti amplificati fluorescenti a lunghezza variabile

FISH (Fluorescence in situ Hybridization)



Ad una sonda progettata si legano molecole di biotina legate al dUTP, in modo da ottenere una marcatura della sonda. La sonda biotinata viene successivamente denaturata, e ibridizzata con un cromosoma che presenta DNA denaturato in situ. Si forma così un ibrido molecolare dato da sonda + DNA cromosomico. Si passa ad una fase di incubazione con avidina fluorescinata, che si lega alle molecole di biotina, permettendo l'osservazione del vetrino al microscopio.

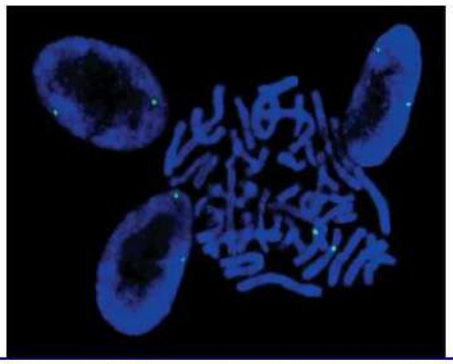


Figura 4.9 Esempio di applicazione della tecnica FISH con sonda centromerica specifica per il cromosoma 8, di cui sono evidenziate due copie. Anche nei nuclei sono presenti due segnali fluorescenti, uno per cromosoma omologo.

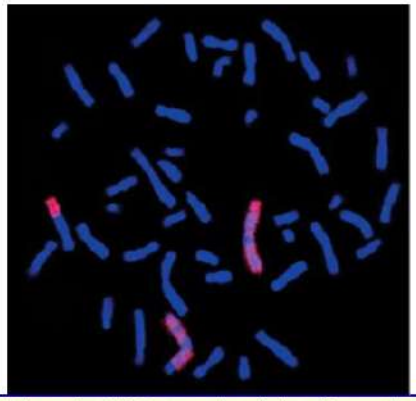


Figura 4.10 Esempio di applicazione della tecnica FISH con sonda painting. Una sonda specifica per il cromosoma 1 ha evidenziato una traslocazione t(1;12).

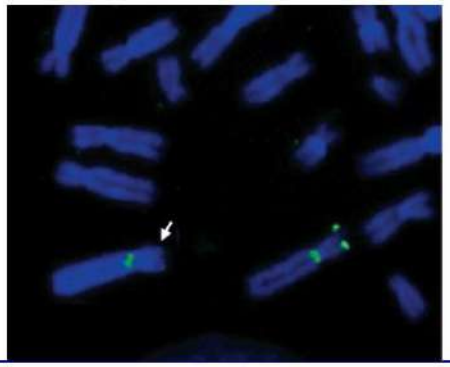


Figura 4.11 Esempio di applicazione della tecnica FISH con sonda *locus*-specifica. In questo caso nel cromosoma 4 normale è presente sia la sonda di repere, che serve a identificare il cromosoma 4 nella metafase, centromerica, sia la sonda subtelomerica specifica per la sindrome di Wolf-Hirschhorn; il cromosoma deleto ha invece solo la sonda centromerica. La delezione osservata è diagnostica per la sindrome di Wolf-Hirschhorn.

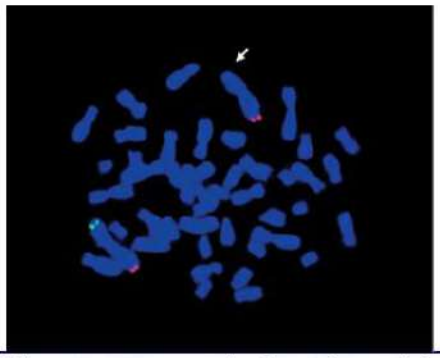


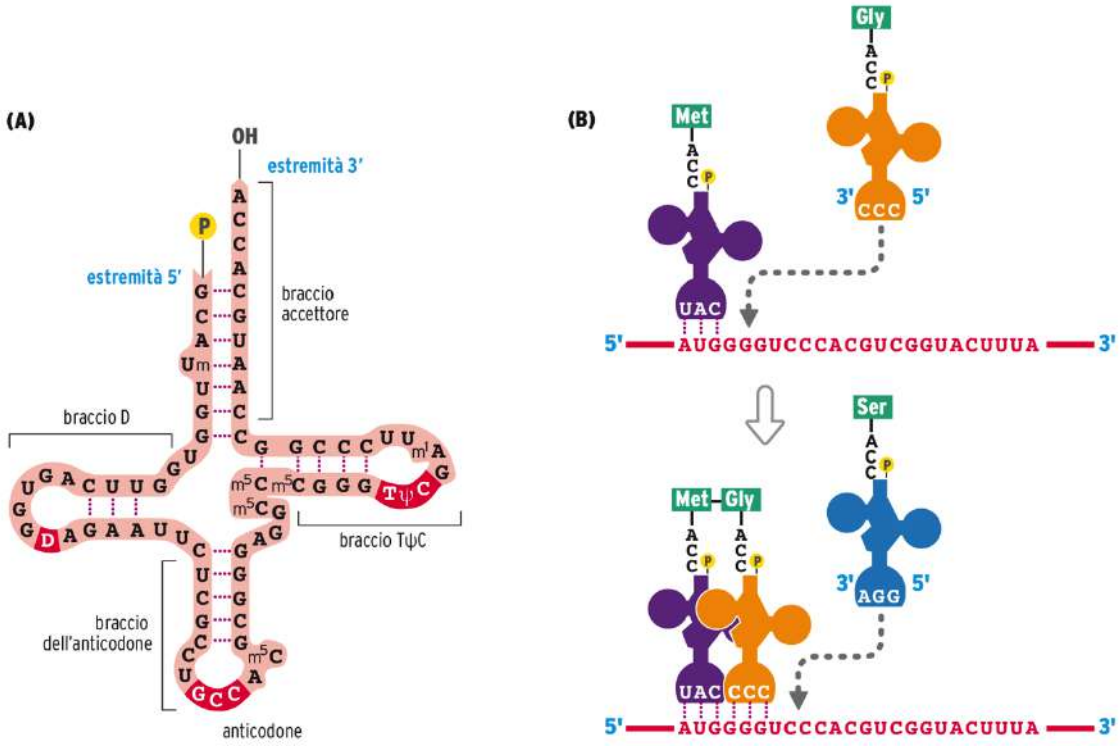
Figura 4.12 Esempio di applicazione della tecnica FISH con sonde telomeriche, specifiche per il braccio corto (verde) e lungo (rosso) del cromosoma 1. La freccia indica delezione del telomero 1p.

I geni per RNA

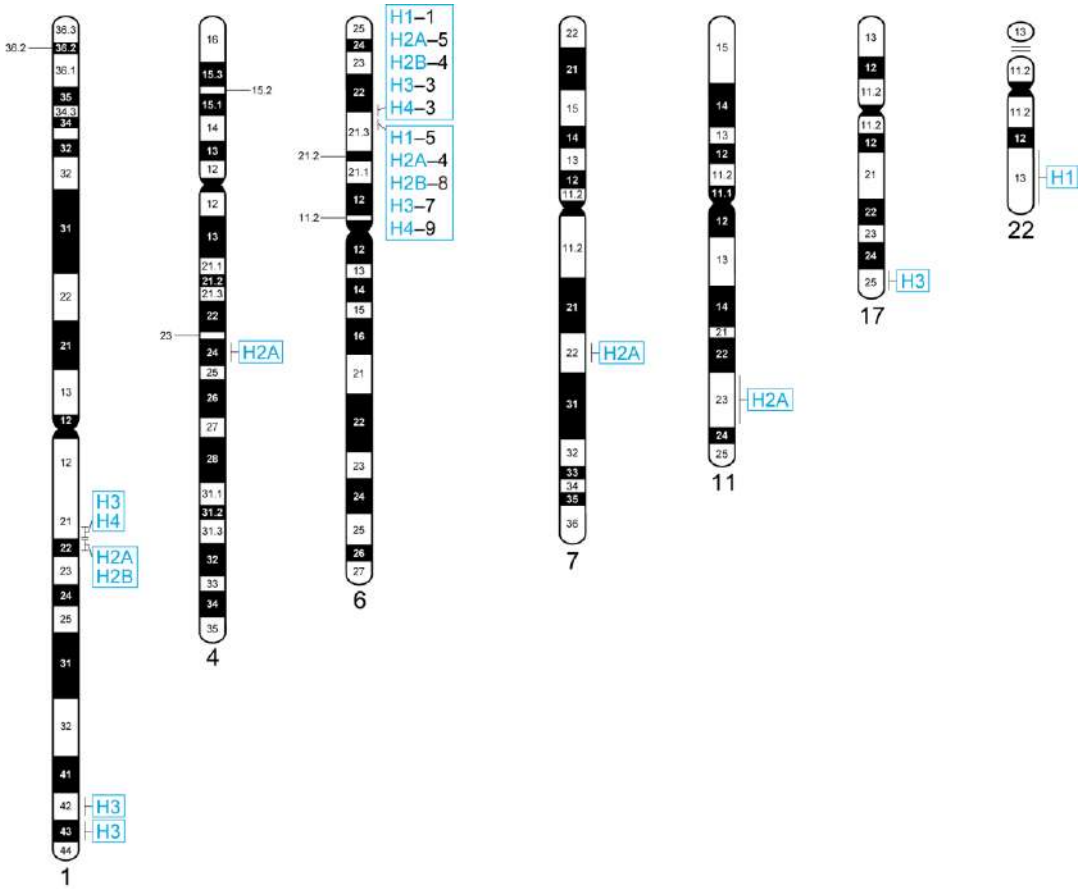
Tra le sequenze altamente conservate, oltre ai geni codificanti per proteine, si annoverano i geni che specificano per RNA o sequenze regolatrici generiche. Il loro numero è comparabile con quello dei geni codificanti per proteine. Alcuni RNA che si formano sono coinvolti nei processi di sintesi ed esportazione delle proteine, come i tRNA, i rRNA, il 7SL RNA.

I geni codificanti per gli rRNA sono costituiti da sequenze ripetute in tandem per 30-40 volte. Queste unità trascrizionali sono lunghe 13 kb e sono separate da sequenze spaziatrici lunghe 27 kb. Queste sequenze di rDNA sono raggruppate in 5 clusters, distribuiti sul braccio corto dei seguenti cromosomi: 13, 14, 15, 21, 22.

Anche i geni per i tRNA sono composti da sequenze ripetute in tandem. L'RNA transfer (o RNA di trasporto), abbreviato in tRNA, è una piccola catena di RNA (costituita da circa 70-90 nucleotidi) che trasferisce un amminoacido specifico di una catena polipeptidica in crescita al sito ribosomiale della sintesi proteica durante la traduzione. Si tratta di circa 500 geni che sono localizzati su tutti i cromosomi ad eccezione dei due cromosomi sessuali. Questa è la struttura di un tRNA e il suo funzionamento come “lettore” dell’mRNA:



Esistono anche geni codificanti per gli istoni, proteine basiche organizzate in ottameri (costituiti da 2 copie degli istoni H2A, H2B, H3 e H4), che si associano strettamente alle molecole di DNA, favorendone il compattamento. Un quinto istone (H1) blocca invece la doppia elica di DNA, come un "fermaglio" sull'ottamero istonico. I geni sono così distribuiti:



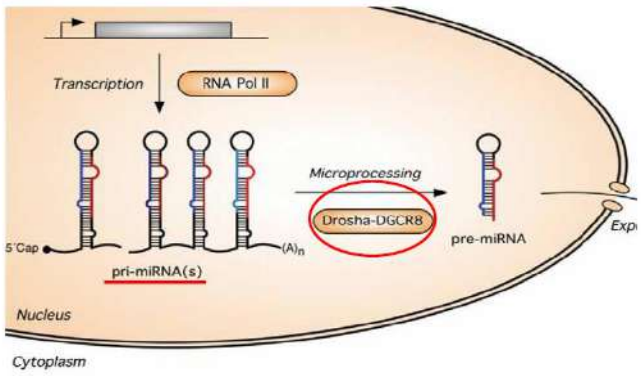
Negli Eucarioti superiori la maggior parte delle sequenze ripetute è DNA non codificante per proteine.

Altri RNA sono invece coinvolti in meccanismi di regolazione genica. Tra questi sono particolarmente importanti i microRNA.

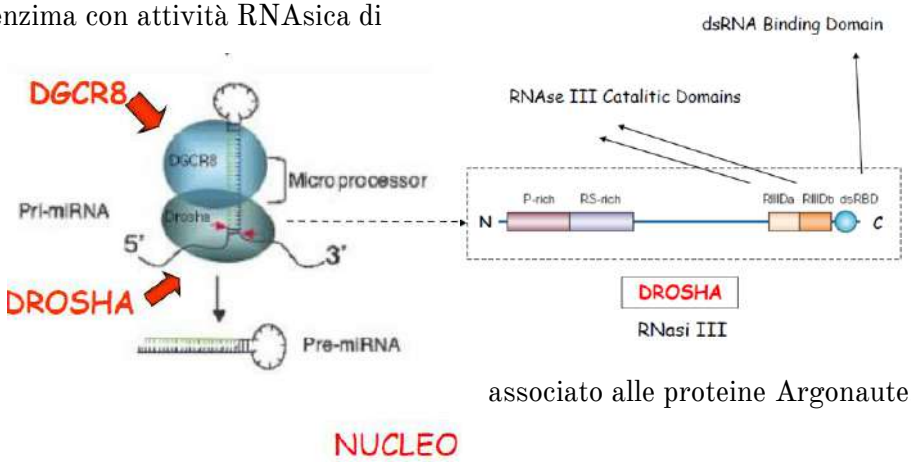
I microRNA

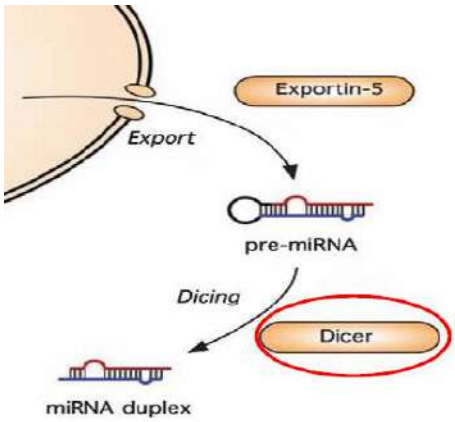
I miRNA compongono una famiglia di ssRNA (single-stranded RNA) di circa 22 nt, che agiscono come regolatori negativi dell'espressione genica a livello post-trascrizionale. I miRNA agiscono, per i 2/3, come unità di trascrizione, come miRNA intronici o esonici. Per un terzo, sono presenti come unità indipendenti. In altri casi, i miRNA sono raggruppati in cluster.

I miRNA presentano un complesso processo di biogenesi che comincia nel nucleo con la trascrizione, a opera della RNA polimerasi II, di RNA più lunghi, denominati microRNA primari (pri-miRNA). I pri-miRNA sono caratterizzati dalla presenza di una struttura secondaria a forcina (stem-loop) costituita da una regione di appaiamento intramolecolare e da due porzioni a singolo filamento. Il pri-miRNA è soggetto, sempre nel nucleo, al taglio da parte di una RNAsi di tipo III denominata DROSHA che, in concerto con la proteina DGCR8, determina la rimozione delle regioni a singolo filamento (5'P e 3' overhang) e la ritenzione unicamente della struttura a forcina, che viene a questo punto denominata pre-miRNA, miRNA precursore. DROSHA, un'endonucleasi, presenta due domini catalitici da RNAsi III e un dominio di legame per il dsRNA. Una via di biogenesi alternativa all'azione del complesso DROSHA-DGCR8 (conservato negli Eucarioti e con un peso molecolare di 650 kDa), e descritta più recentemente, prevede l'origine del pre-miRNA direttamente dal processo di splicing di regioni introniche. In quest'ultimo caso, i miRNA che si formeranno prenderanno il nome di mirtroni.

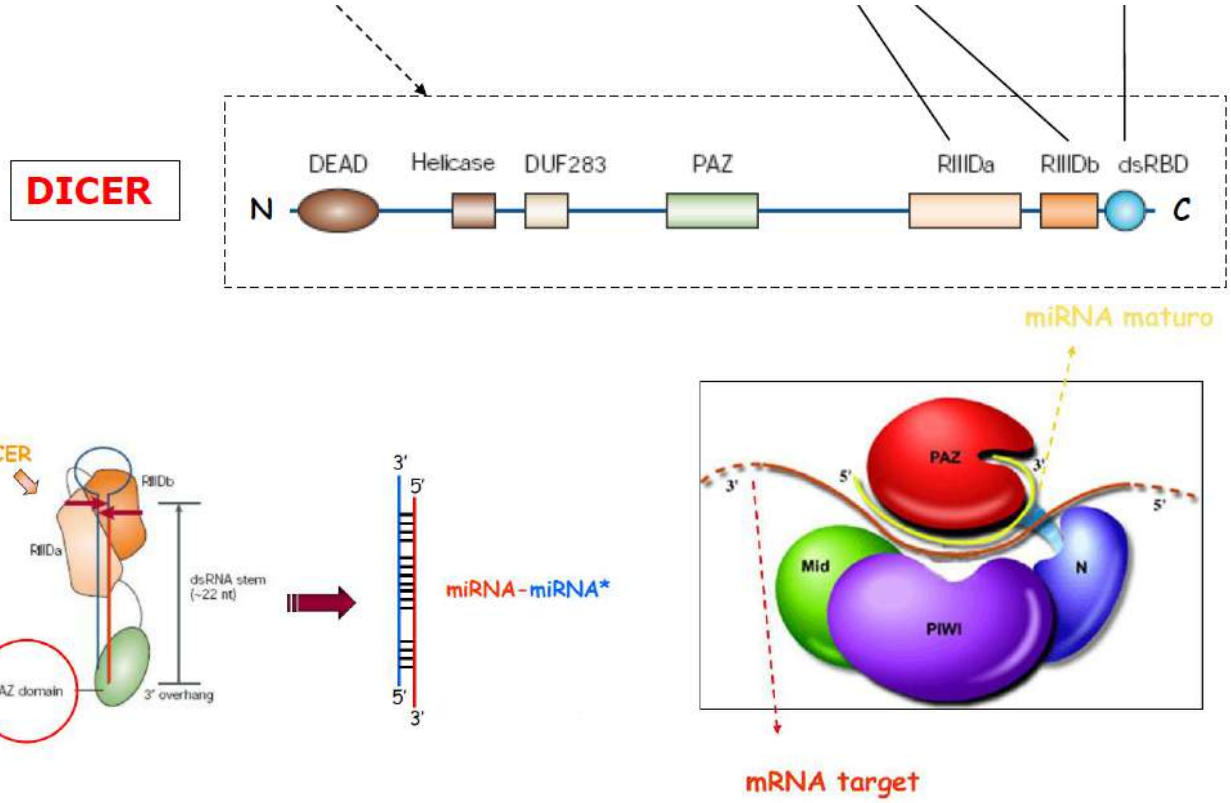


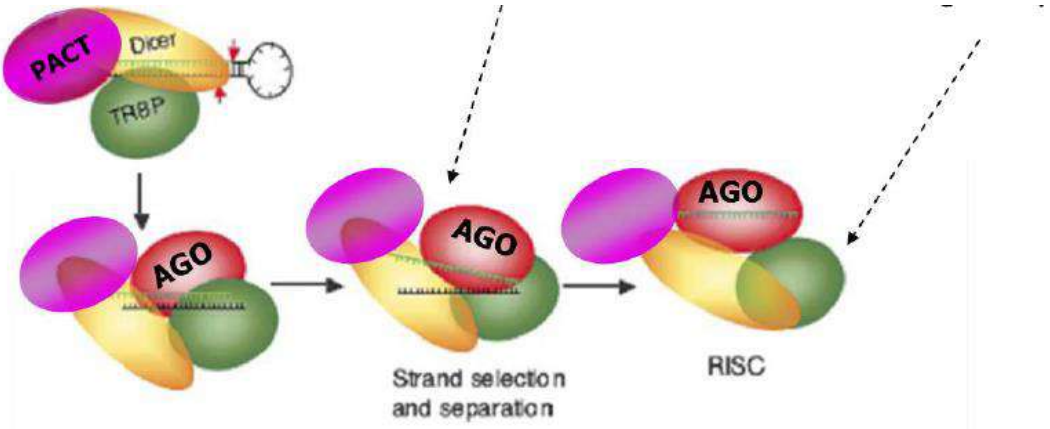
Il pre-miRNA viene esportato nel citoplasma dall'esportina-5, dove viene processato da un altro enzima con attività RNAsica di tipo III (Dicer), insieme a TRBP e PACT, che rimuove l'ansa della forcina e produce un miRNA maturo a doppio filamento con estremità sporgenti. Dicer riconosce le estremità sporgenti ed è associato alle proteine Argonaute.



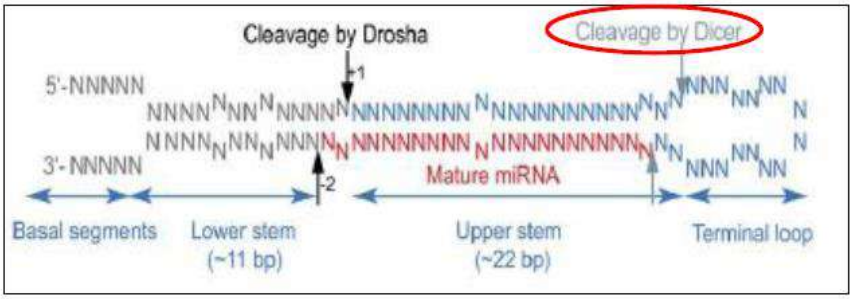


Possiede gli stessi domini di DROSHA. Il dominio PAZ è importante perché facilità il riconoscimento tra molecole di RNA. I miRNA maturi si legano al complesso di silenziamento indotto da RNA (RISC, RNA-Induced Silencing Complex) costituito dalla proteina con attività endoribonucleasica denominata Argonauta (Ago), che svolge il doppio filamento, come un'elicasi. La proteina Argonauta è composta da un dominio PAZ N-terminale e uno PIWI C-terminale.



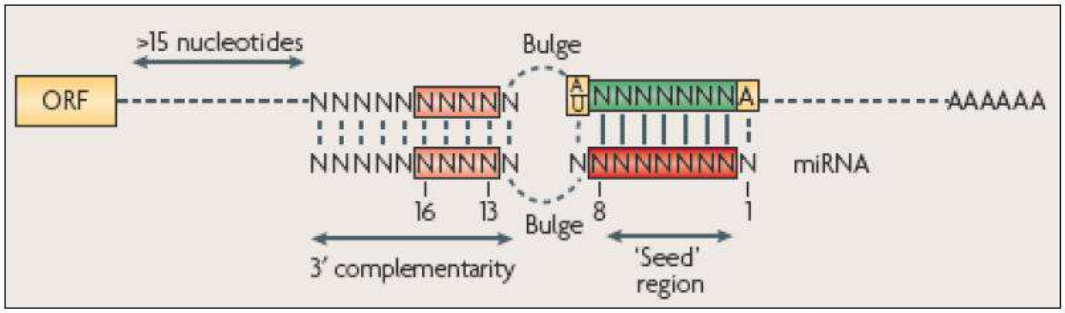


Inizialmente si riteneva che in seguito all'azione del complesso RISC solo uno dei due filamenti (il cosiddetto filamento "guida") agisse come miRNA maturo finale, mentre quello complementare ("passeggero") venisse completamente degradato. Tuttavia, è ora accertato che entrambi i filamenti

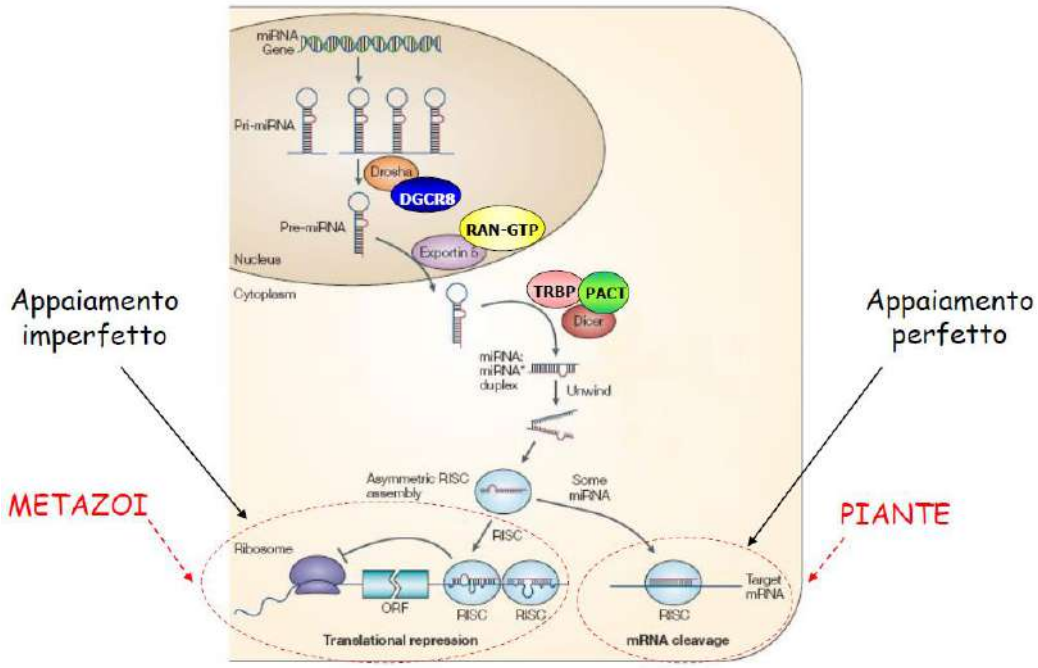


sono in grado di funzionare come miRNA maturi, anche se in generale uno dei due presenta una stabilità maggiore e assume un ruolo predominante.

Il complesso miRNA-RISC è, a questo punto, in grado di legare sequenze di mRNA bersaglio che presentano regioni di parziale complementarità con il miRNA stesso, regioni localizzate prevalentemente nelle loro regioni 3' non tradotte. All'interno della regione di legame, tuttavia, esiste una porzione di perfetta complementarità di sequenza tra il miRNA e il gene bersaglio che corrisponde a una regione di 7-8 nucleotidi circa all'estremità 5' del miRNA, denominata "**seed**".



Se si verifica un appaiamento perfetto, come nelle piante, l'mRNA target va incontro a degradazione, mentre nei Metazoi, dove l'appaiamento è imperfetto, si verifica una repressione della traduzione.

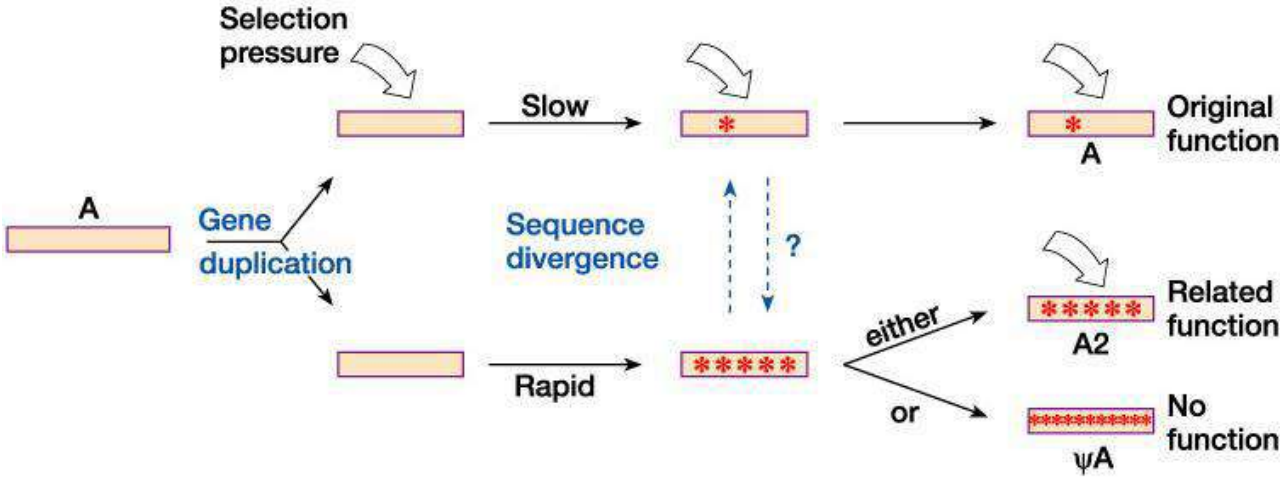


Altri geni codificanti RNA sono i lncRNA (long non-coding RNA), e gli RNA endogeni non competitivi, come gli RNA circolari.

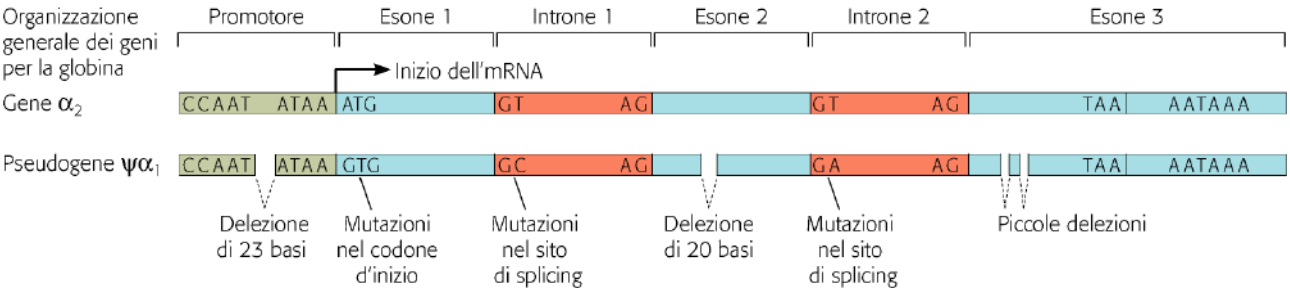
Gli pseudogeni

Nel nostro genoma si trova inoltre un alto numero (circa 14 600 secondo le stime più recenti) di copie non funzionali di geni detti **pseudogeni**, che si sono originati con due meccanismi diversi: per duplicazione o per retrotrascrizione e successivo inserimento in altra parte del genoma.

Immaginiamo la situazione di un gene ancestrale A, che, in seguito a fenomeni come il crossing-over ineguale, subisce un processo di duplicazione, producendo due copie dello stesso gene. Una delle due copie, sottoposta a pressione selettiva, accumula, lentamente, soltanto qualche mutazione sporadica comunque non inattivante, che non ne ostacola la funzione, anzi, la porta a occupare il ruolo che apparteneva al gene ancestrale, ma in maniera più specifica. L'altra copia, invece, attira, perché le è consentito, su di sé, in breve tempo, una serie di mutazioni nucleotidiche che contribuiscono alla trasformazione della sequenza originale, che, dunque, si perde.

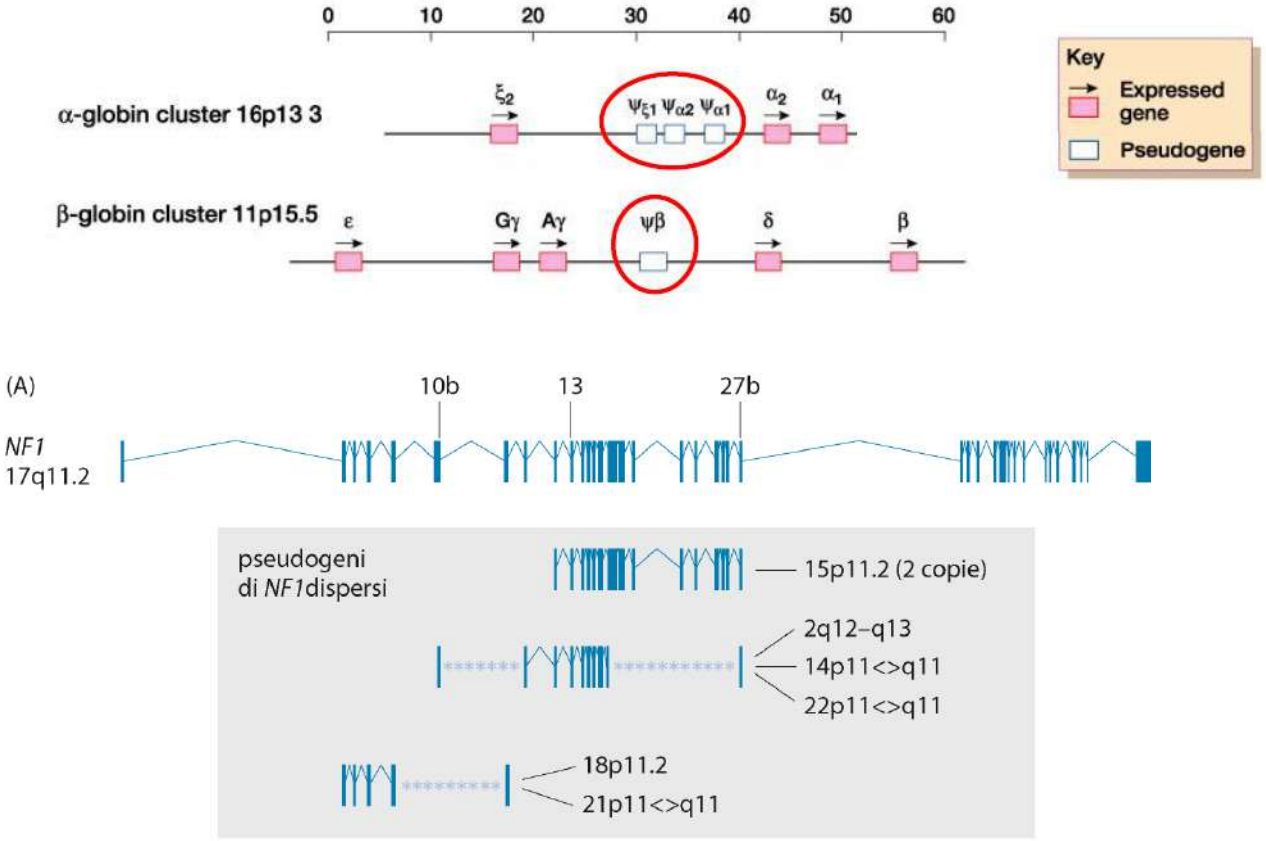


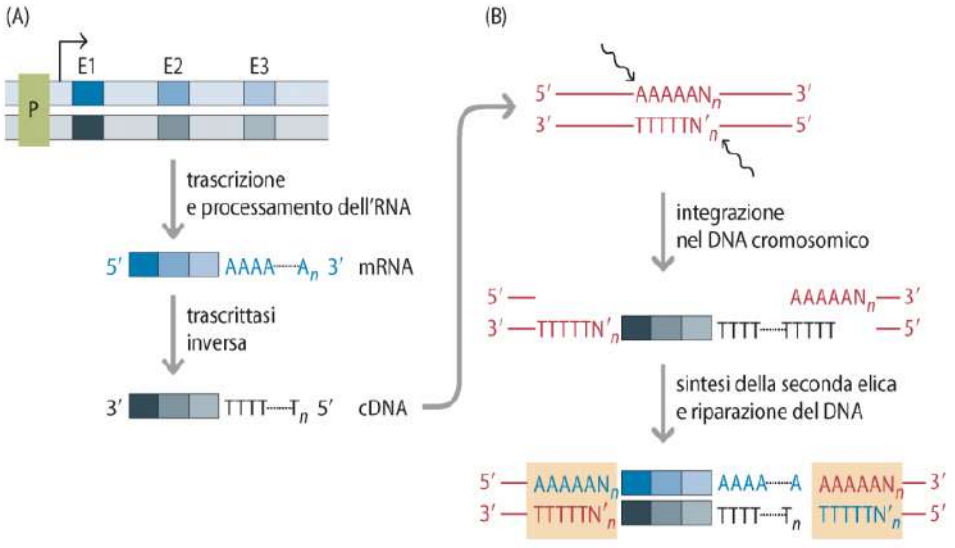
Il nuovo gene, un omologo di quello funzionale, può o svolgere una funzione alternativa e correlata al gene funzionale, oppure perdere completamente ogni funzione (se le mutazioni acquisite sono inattivanti) e diventare uno pseudogene non processato, come uno dei geni del cluster beta-globinico, oppure lo pseudogene psialfa2 del cromosoma 16, nel gene che codifica per la alfa-globina.



Gli pseudogeni non processati, quindi presentano una struttura completa di esoni e introni, e si trovano vicino agli pseudogeni omologhi funzionanti. Derivano da duplicazioni. Possono essere "completi" (come nel caso dell'alfa e della beta globina, cromosomi 16 e 11) o "parziali", di cui

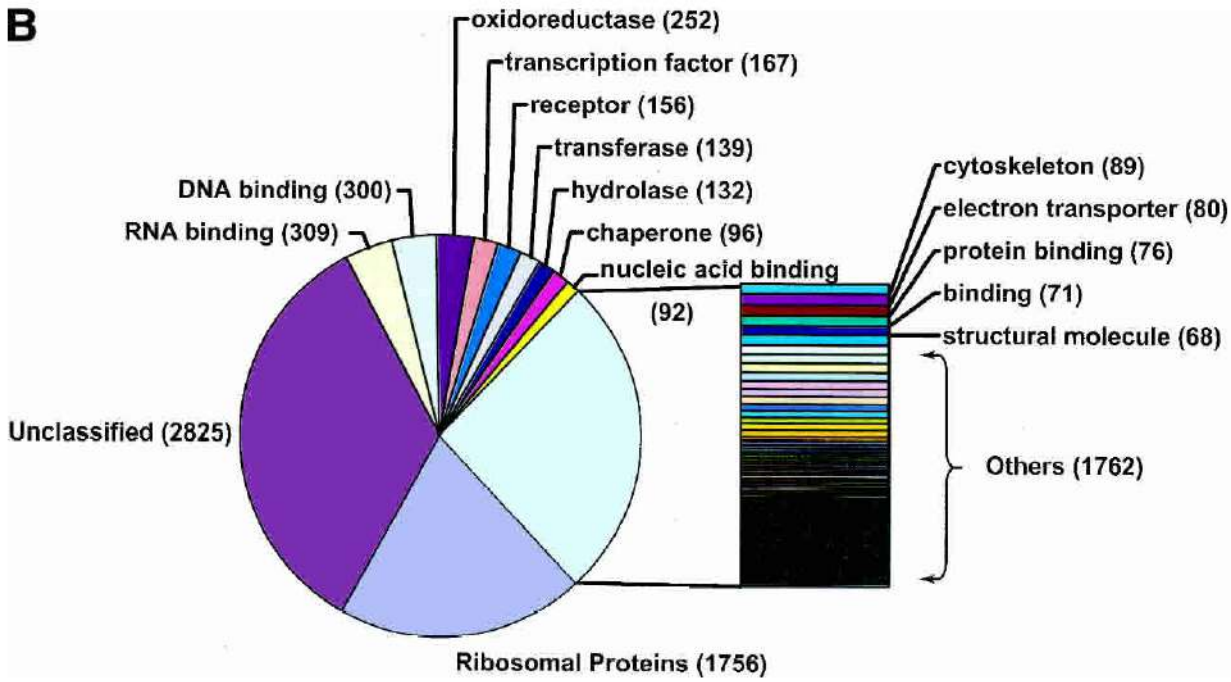
rappresenta un esempio il gene NF1 (per la neurofibromina 1), i cui pseudogeni non processati sono dispersi nel genoma su differenti cromosomi.





Gli pseudogeni processati, invece, presentano soltanto esoni e derivano dalla retrotrasposizione. Sono geni che sono stati trascritti e il cui RNA messaggero è stato processato, cioè sottoposto a splicing. Ad opera della trascrittasi inversa cellulare dall'RNA messaggero del gene ancestrale è stato sintetizzato un singolo filamento di DNA complementare antisenso, che si è andato ad inserire, casualmente in un altro sito cromosomico. E' stata poi sintetizzata una seconda elica complementare a questo cDNA ed è stato riparato il punto di inserimento. Gli pseudogeni processati, derivando da RNA messaggeri, possono avere una coda di poliA, ma sono privi di promotori e quindi sono solo raramente trascritti.

Nel 2003 erano stati identificati approssimativamente 8000 pseudogeni processati, dei quali il 94% era completo. Questo è un grafico degli pseudogeni processati attualmente classificati:



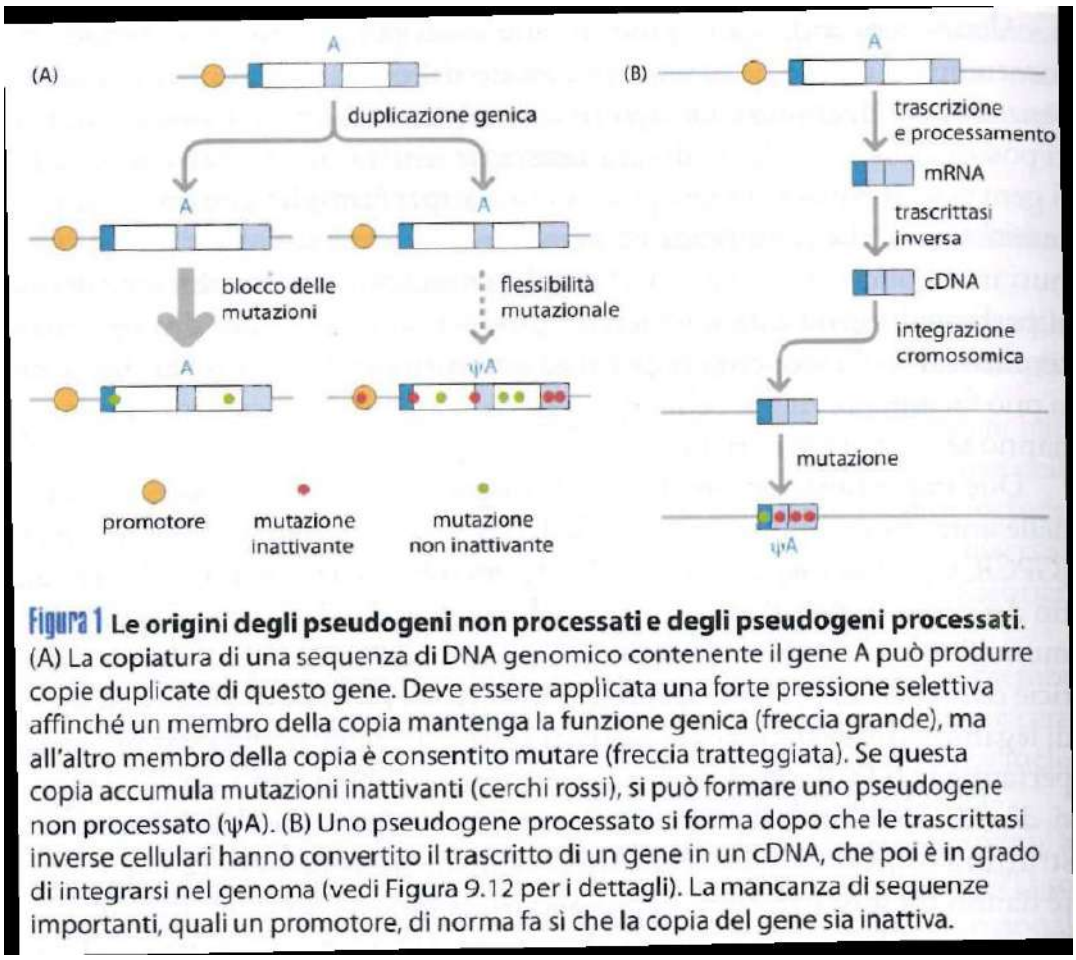


Figura 1 Le origini degli pseudogeni non processati e degli pseudogeni processati. (A) La copiatura di una sequenza di DNA genomico contenente il gene A può produrre copie duplicate di questo gene. Deve essere applicata una forte pressione selettiva affinché un membro della copia mantenga la funzione genica (freccia grande), ma all'altro membro della copia è consentito mutare (freccia tratteggiata). Se questa copia accumula mutazioni inattivanti (cerchi rossi), si può formare uno pseudogene non processato (ψA). (B) Uno pseudogene processato si forma dopo che le trascrittasi inverse cellulari hanno convertito il trascritto di un gene in un cDNA, che poi è in grado di integrarsi nel genoma (vedi Figura 9.12 per i dettagli). La mancanza di sequenze importanti, quali un promotore, di norma fa sì che la copia del gene sia inattiva.

Schema riassuntivo

Se questo evento si verifica nella linea germinale e il trascritto si integra casualmente nel genoma, molto probabilmente non sarà in grado di esprimersi (mancherà infatti di un promotore) e, dopo l'acquisizione di mutazioni dannose, subirà una degenerazione a retropseudogene. Se tuttavia il trascritto si inserisce accanto a un promotore endogeno, potrebbe essere espresso. Molto raramente, e in base alla necessità, il gene espresso può essere utile e quindi può essere conservato come retrogene funzionale.

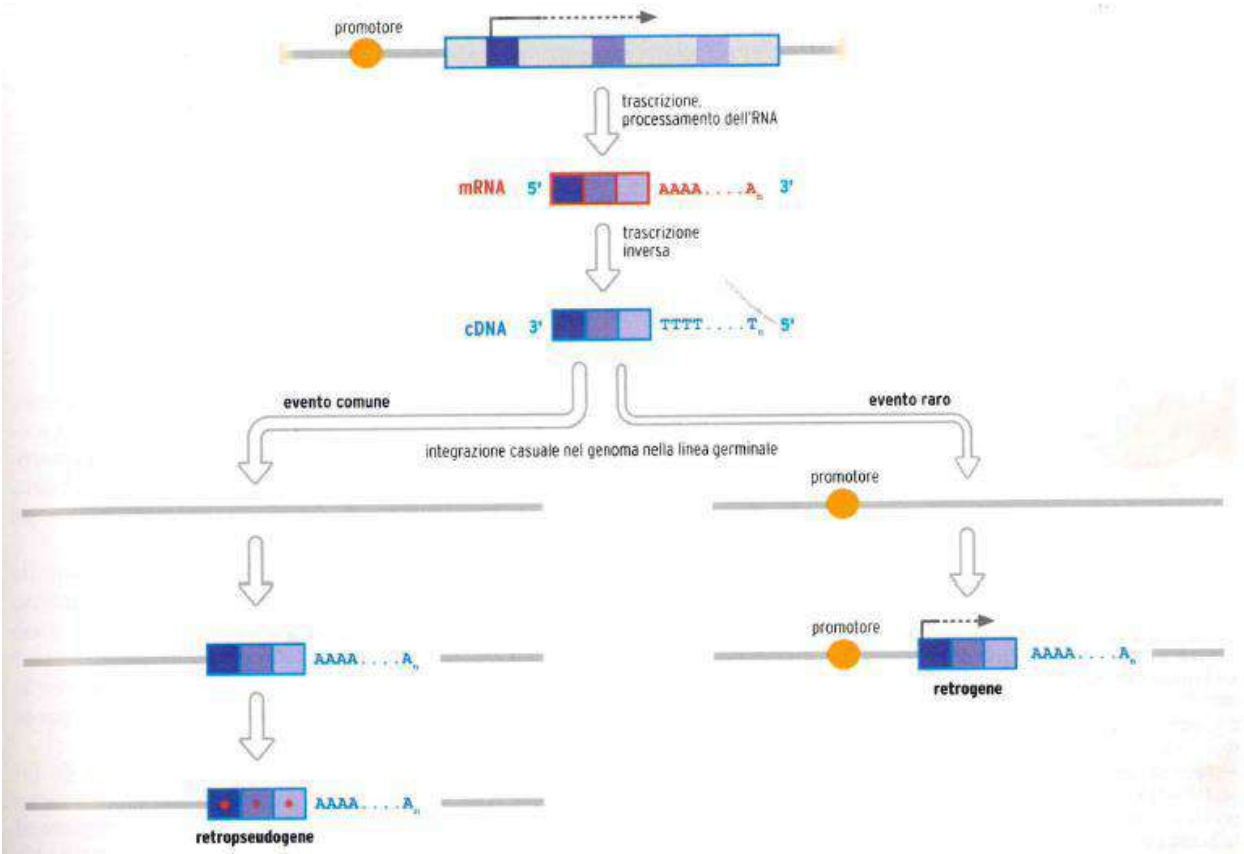


Figura 1 ▲ I retrogeni e i retropseudogeni derivati dalla trascrizione inversa da trascritti di RNA. In questo esempio un gene che codifica proteine con tre esoni è trascritto da un promotore a monte e gli introni sono tagliati dal trascritto per produrre un mRNA. L'mRNA può poi essere convertito ovviamente in un cDNA antisenso a singolo filamento utilizzando una funzione della trascrittasi inversa cellulare. Se ciò si verifica nella linea germinale e il trascritto si integra casualmente nel genoma, molto probabilmente non sarà in grado di esprimersi (mancherà infatti di un promotore) e, dopo l'acquisizione di mutazioni dannose (asterischi rossi in basso a sinistra), subirà una degenerazione a retropseudogene. Se tuttavia il trascritto si inserisce accanto a un promotore endogeno potrebbe essere espresso. Molto raramente, e in base alla necessità, il gene espresso può essere utile e quindi può essere conservato come retrogene funzionale (in basso a destra).

Tabella 9.11: **Esempi di retrogeni umani senza introni e dei loro omologhi parentali contenenti introni**
(per maggiori informazioni, si veda http://exppc01.uni-muenster.de/expath/frames.htm)

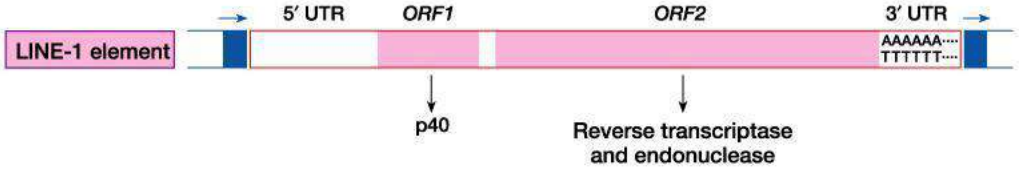
Retrogene	Omologo contenente introni	Prodotto
GK2 in 4q13	*GK1* in Xp21.3	Glicerolo chinasi
PDHA2 in 4q22-q23	*PDHA1* in Xp22	Piruvato deidrogenasi
PGK2 in 6p12.3	*PGK* in Xq13	Fosfoglicerato chinasi
TAF1L in 9p13.3	*TAF1* in Xq13.1	Fattore associato alla proteina che si lega a TATA box, 250 kDA
MYCL1 in 1p34.2	*MYCL2* in Xq22-q23	Omologo dell'oncogene v-myc
GLUD1 in 10q23.3	*GLUD2* in Xq25	Glutammato deidrogenasi
SNAIL1L1 in 2q33-q37	*SNAIL1* in 20q13	Regolatore dello sviluppo associato a snail

Le informazioni che emergono dimostrano che molti retrogeni si sono originati da progenitori genici legati al cromosoma X, sono autosomici e sono specificatamente espressi nei testicoli. Durante la meiosi maschile nei Mammiferi, entrambi i cromosomi sessuali sono inattivi dal punto di vista trascrizionale e segregati nel cosiddetto corpo "XY", mentre gli autosomi sono attivi. Dal momento che il cromosoma X è altamente conservato nei Mammiferi, l'evoluzione di retrogeni specifici è stato ipotizzato che compensi il silenziamento dei loro geni progenitori legati all'X durante la spermatogenesi.

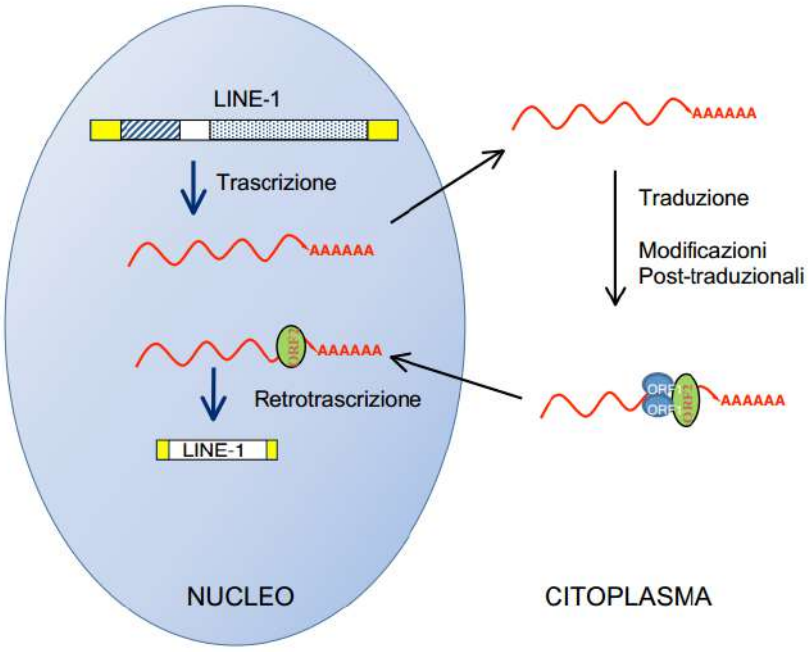
Esiste anche un database nel quale sono stati raccolti tutti i retrogeni noti in vari organismi animali. Nel caso dell'uomo sono stati censite 4927 retrocopie, 914 delle quali hanno mantenuto le proprie ORF e 615 sono espresse.

LINE e SINE

Le ripetizioni trasposoniche fanno parte delle sequenze scarsamente conservate, altamente ripetitive non in tandem. Tra queste vengono annoverate i LINE (Long Interspered Nucleotide Element), ripetizioni retrotrasposoniche, di 6-8 kb.



Effettuano il jumping e compongono il 21% del genoma. Sono elementi trasponibili autonomi, in particolare L1 si traspone attivamente. Il DNA LINE è composto da due ORF, la prima codifica per una proteina, p40, la seconda per la trascrittasi inversa e la endonucleasi.

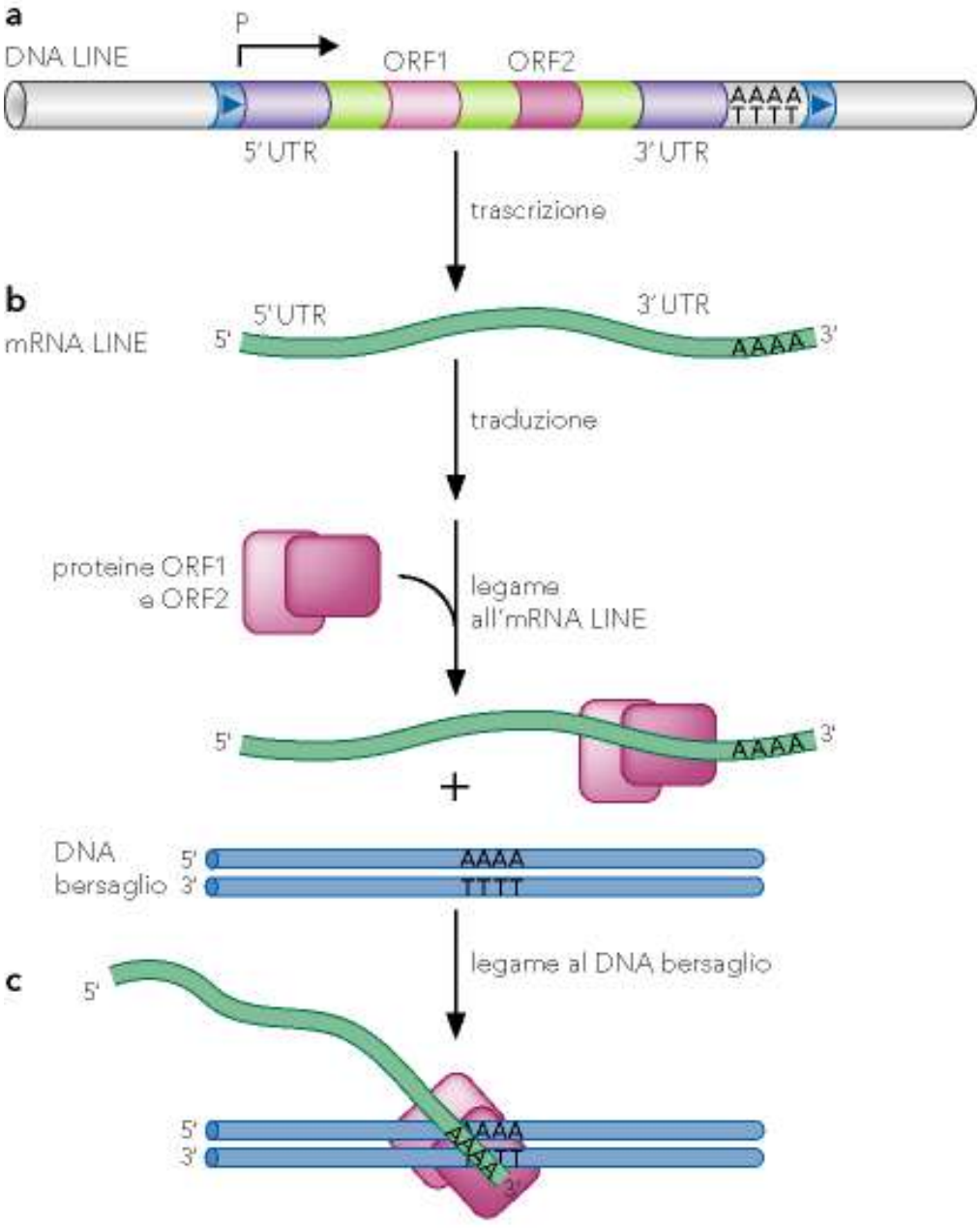


L'elemento LINE-1 è trascritto nel nucleo, l'RNA messaggero poi viene trasportato nel citoplasma. ORF1 e ORF2 sono tradotte, avvengono modifiche post-traduzionali e si associano al loro stesso RNA messaggero. Il complesso ribonucleoproteico viene rilocalizzato nel nucleo e retro trascritto.

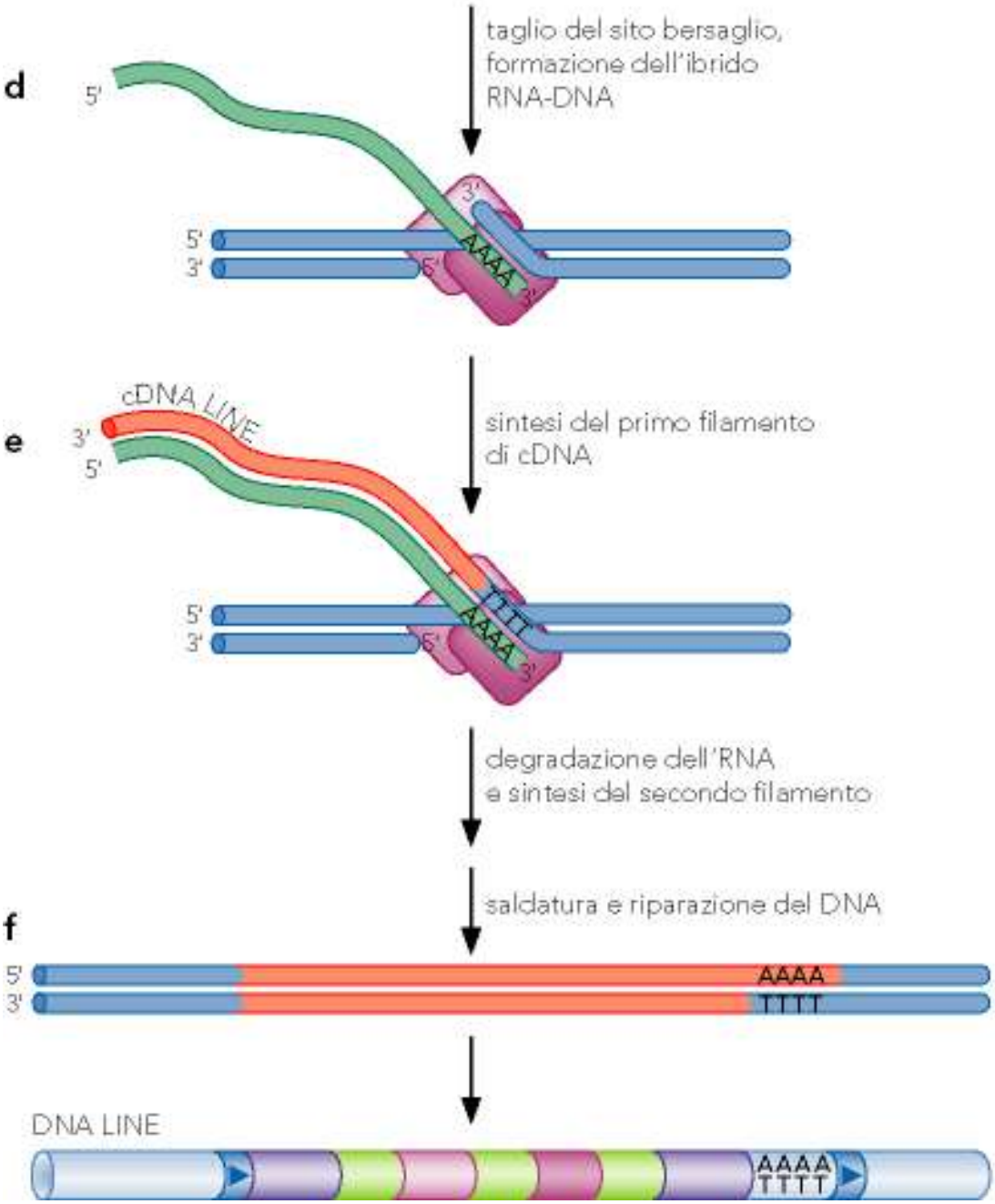
Fig. 3 **Meccanismo di retrotrasposizione**. L'elemento LINE-1 è trascritto, l' mRNA viene trasportato nel citoplasma. ORF1 e ORF2 sono tradotte, si associano loro stesso mRNA , il complesso ribonucleoproteico è rilocalizzato nel nucleo e retrotrascritto.

TPRT (Target-Primed Reverse Transcription) è il nome del meccanismo con il quale i LINE si inseriscono nel DNA. Esso è composto dai seguenti passaggi:

1. Trascrizione
2. Traduzione
3. Legame all'RNA messaggero LINE
4. Formazione di un legame con il DNA bersaglio

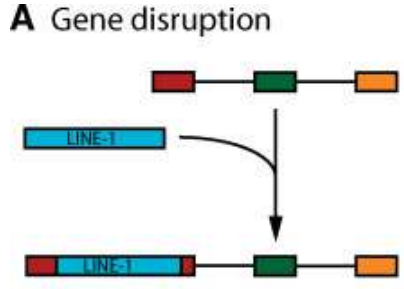


5. Taglio del sito bersaglio
6. Formazione dell'ibrido DNA-RNA
7. Sintesi del primo filamento di cDNA
8. Degradazione dell'RNA e sintesi del secondo filamento
9. Saldatura e riparazione del DNA
10. Formazione nuovo DNA LINE



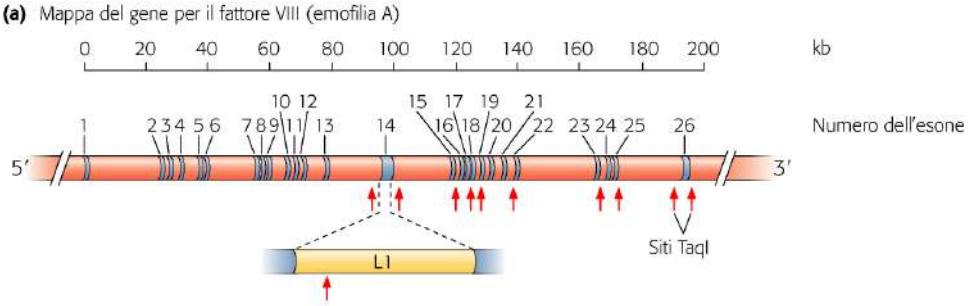
1 processo su 100 va a completamento.

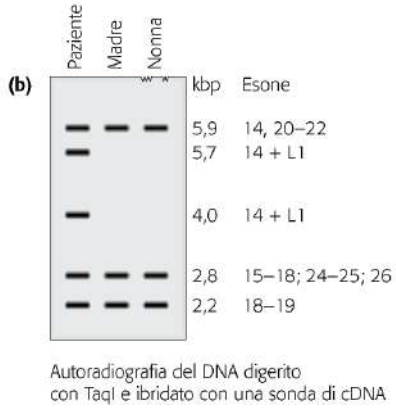
Delle 6.000 LINE-1 a lunghezza completa, 60-100 sono ancora capaci di trasporsi.



Diversi sono i danni genomici indotti da L1. Un caso è costituito dall'interruzione dei geni per effetto di L1. Ciò si verifica in circa 1 su 600 mutazioni osservate. In uno studio pubblicato su "Nature" nel 1988, vennero riportati gli effetti patologici (con la comparsa dell'emofilia di tipo A) di inserzioni di elementi L1 (collegate all'attivazione di un segnale pA debole) all'interno dell'esone 14 del gene del fattore VIII in due di 240 pazienti non imparentati affetti da questa malattia. In elettroforesi del

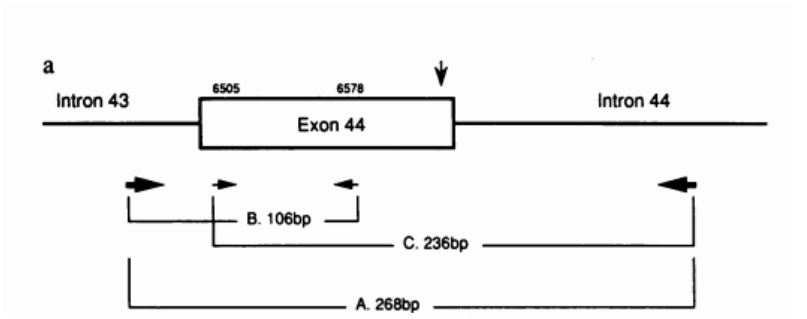
paziente rispetto a non malati appare una banda di lunghezza inaspettata, data dalla somma di E14+L1.

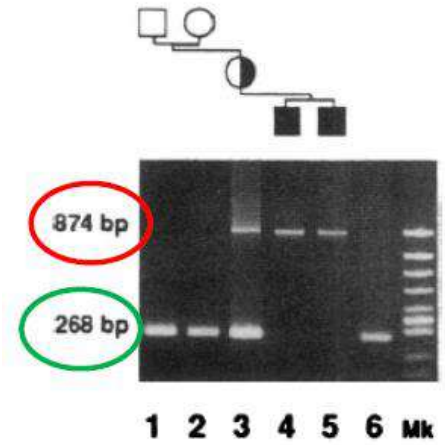




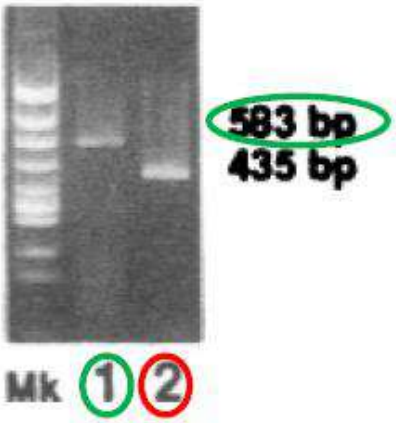
Autoradiografia del DNA digerito con TaqI e ibridato con una sonda di cDNA

La seconda prova della retrotrasposizione dell'elemento L1, che venne trovato inserito all'interno del gene della distrofina di un paziente, provocando la distrofia muscolare di Duchenne (DMD), nel 1993, presso l'ospedale Yakumo di Hokkaido.

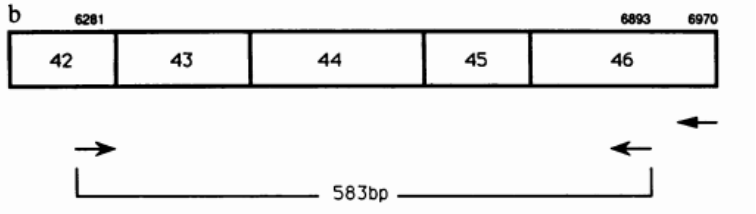




Venne amplificato l'esone 44 del gene della distrofina, estratto da 5 membri della famiglia a cui appartenevano due fratelli affetti da DMD. I frammenti amplificati dai pazienti migravano più lentamente di quello proveniente da un individuo sano. In base alla sua mobilità, la lunghezza del frammento amplificato doveva essere di circa 900 bp, che significa, all'incirca, un incremento di 600 bp rispetto al frammento normale (268 bp). La segregazione dell'inserto venne studiata nei membri della famiglia. La madre aveva sia il frammento normale sia quello più esteso ed è perciò una portatrice. Il solo prodotto normale amplificato venne ottenuto dai nonni materni.



Dall'amplificazione del cDNA preparato dall'RNA messaggero di un linfocita. Il prodotto amplificato comprendente gli esoni da 42 a 46 venne ottenuto con una PCR semi-nested. Il frammento atteso corrispondente a 583 bp venne ottenuto dal controllo.



Al contrario, né il frammento con taglia normale né il frammento corrispondente alla mutazione dovuta all'inserzione furono amplificati dal cDNA del paziente. Invece, il frammento del paziente, amplificato, migrò più velocemente rispetto a quello di un individuo normale.

In base alla sua mobilità, la lunghezza del frammento amplificato del paziente venne calcolato essere circa 435 bp, significando una riduzione di circa 150 bp in confronto al frammento w-t.

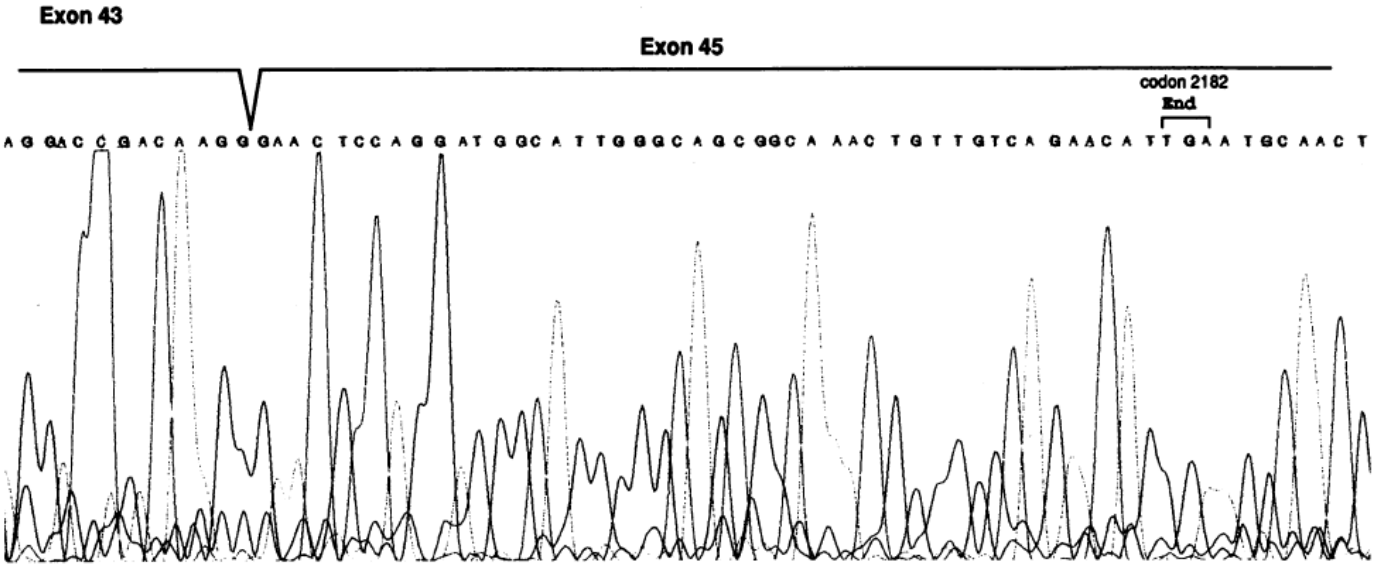
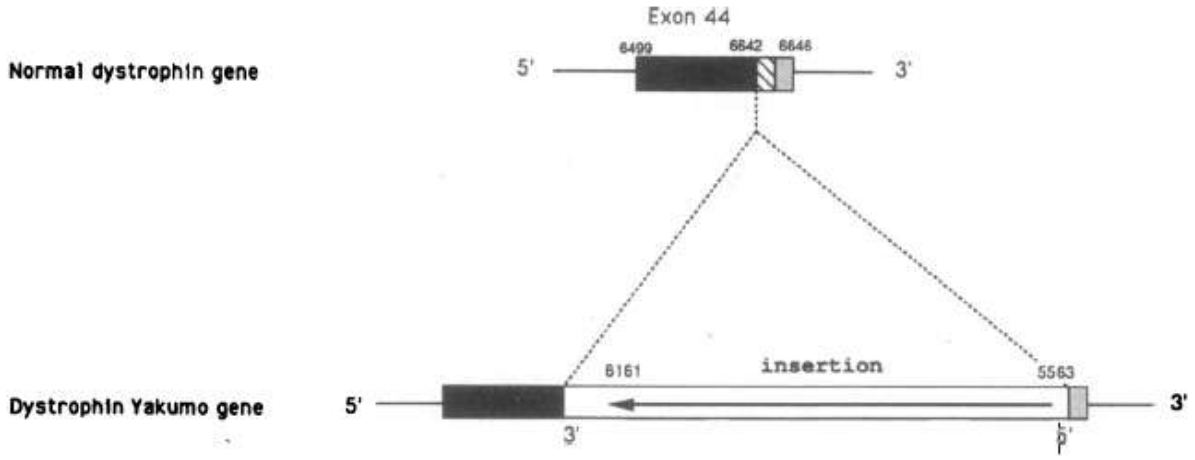


Figure 6. Sequence of the amplified product from patient's cDNA. Sequences of the region joining exons 43 to 45 are presented. The sequence of the 3′ end of exon 43 (5′-ACAAGG-3′) is joined directly to the 5′ end of exon 45 (5′-GAACTCC-3′). Because of this exon skipping, a stop codon (TGA) appears at codon 2182.

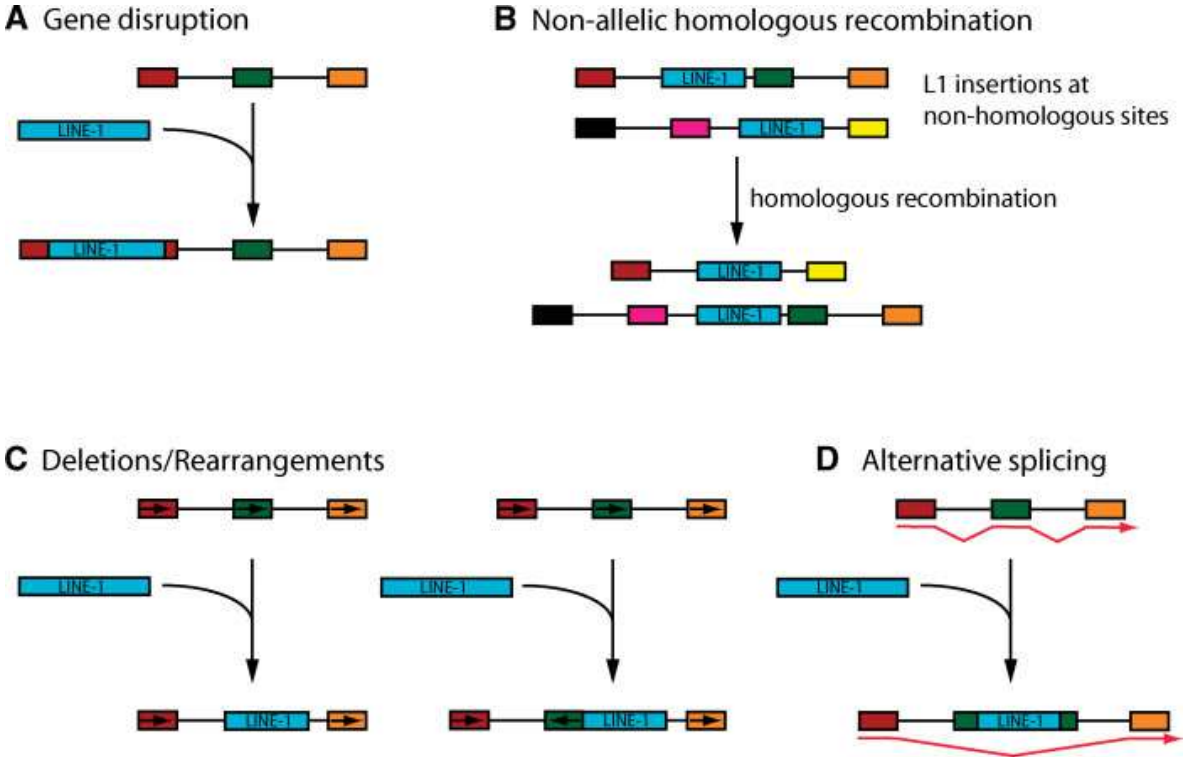
L'exon skipping che si verificava a livello dell'esone 44 provocava l'insorgenza di un codone di stop (UGA) al codone 2182.

Il gene della distrofina "giapponese" presentava l'inserzione, capovolta, di L1.

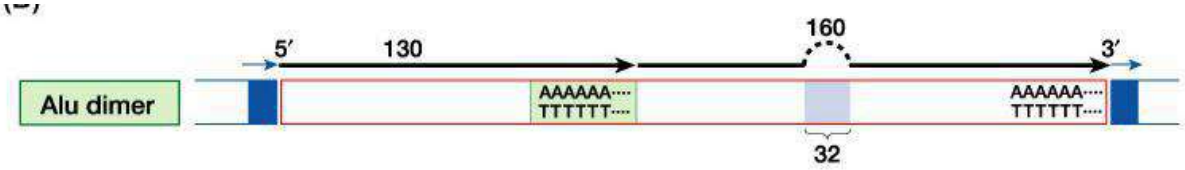


Una su 600 è in realtà una sottostima, metà di quelle trovate sono X-linked.

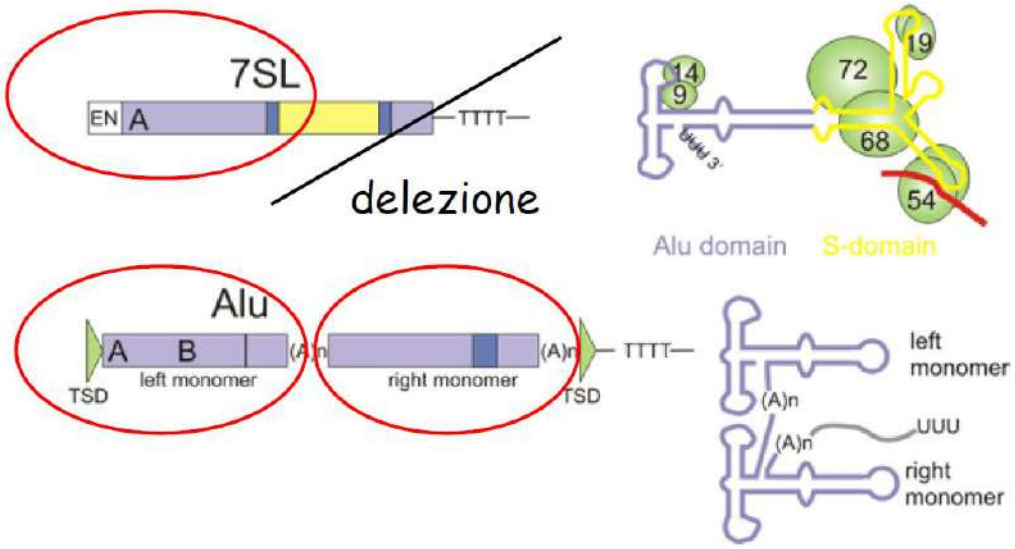
L'inserimento di LINE provoca: interruzione dei geni, ricombinazione non allelica dei cromosomi omologhi, delezioni e riarrangiamenti, splicing alternativo.



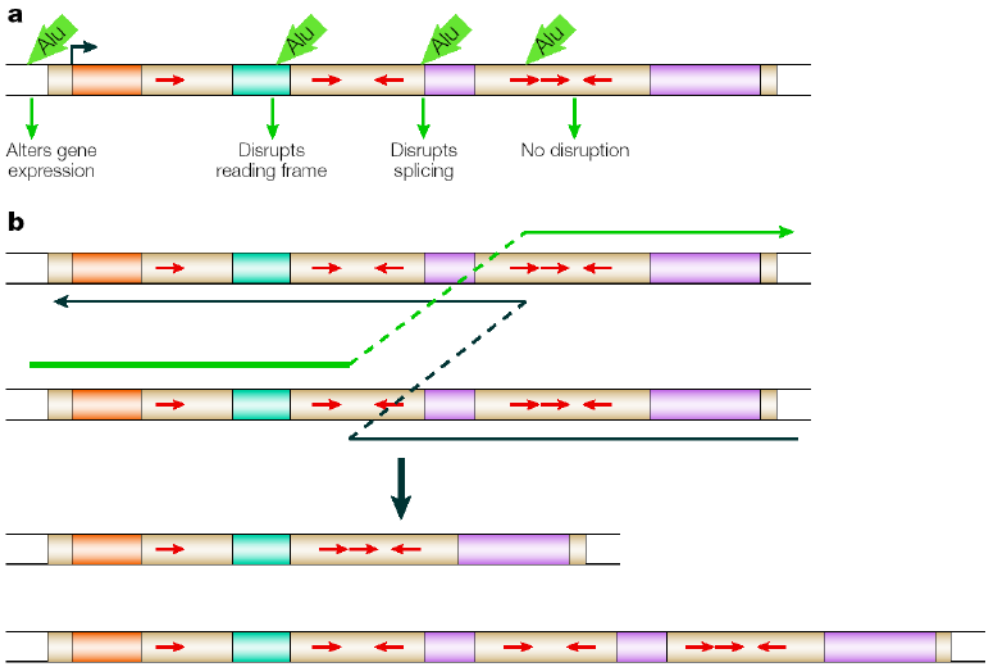
Altre ripetizioni retrotrasposoniche sono costituite dai SINE. Essi sono abbondanti nelle regioni attivamente trascritte. Non codificano per nessuna proteina. Si trova un Alu ogni 3 kb. Una sequenza Alu completa è lunga 280 bp. La loro trasposizione è LINE-dipendente. Sono formati da dimeri. Alu è il nome dell'endonucleasi che riconosce queste sequenze come sito di taglio.



Queste sequenze derivano dall'RNA 7SL, un componente del signal recognition particle, implicato nella veicolazione delle proteine del reticolo endoplasmatico. La struttura primaria dell'RNA 7SL consiste in una sequenza Alu interrotta da una sequenza di 155 bp che è unica per il RNA 7SL. Si ha un monomero di sinistra e un monomero di destra.

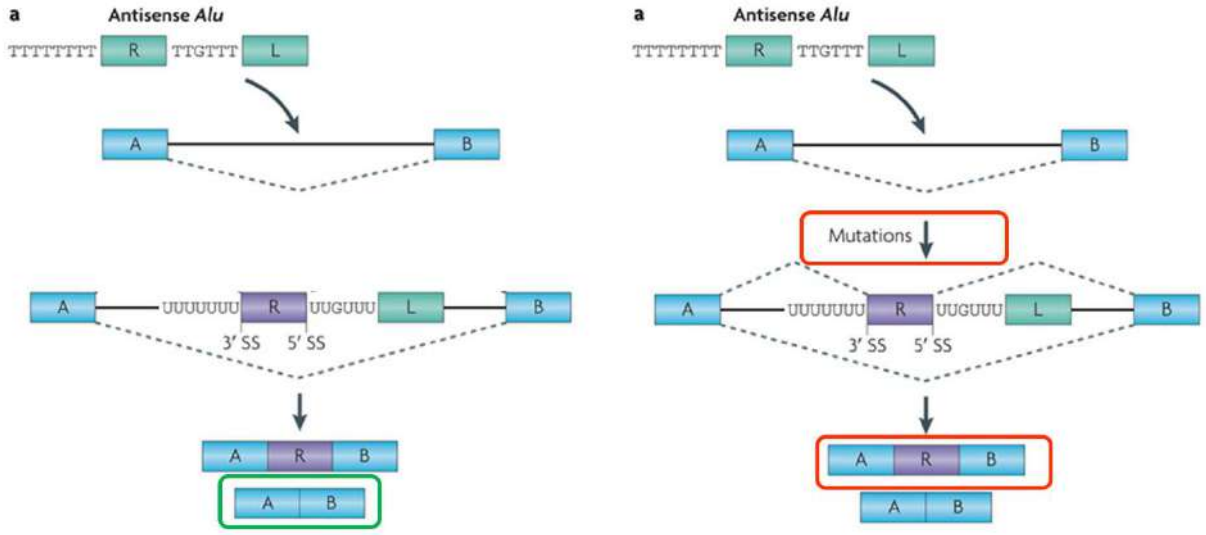


Le sequenze Alu possono alterare l'espressione genica, guastare la cornice di lettura oppure il meccanismo di splicing. Uno studio del 1996 dimostrò che una delle mutazioni che coinvolgevano il gene BRCA2 nell'insorgenza del cancro primario al seno avveniva nella linea germinale, in seguito all'inserzione di un elemento Alu nell'esone 22, che provocava uno splicing alternativo che saltava l'esone 22. Altre patologie sono provocate dall'integrazione casuale di Alu (neurofibromatosi, emofilia,...) o crossing-over ineguale (diabete di tipo II, sindrome di Lesch-Nyhan, malattia di Tay-Sachs, alfa-thalassemia e alcuni tipi di cancro).

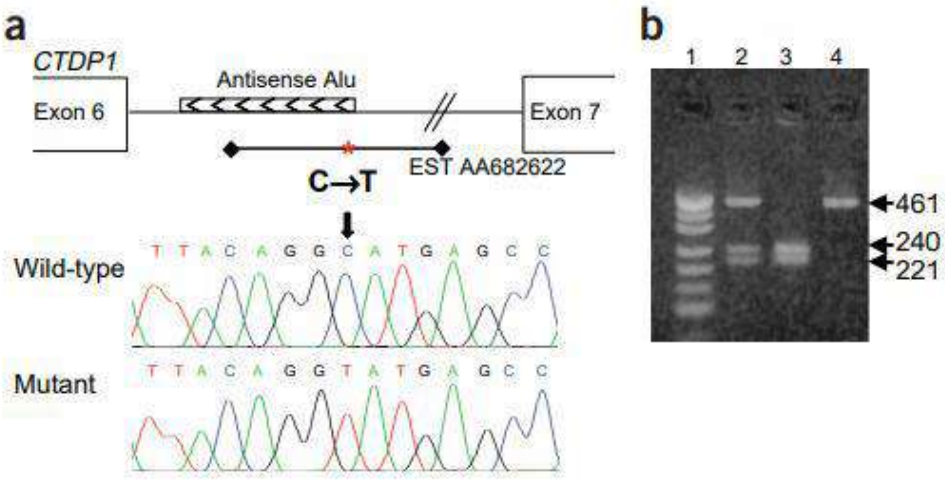


Si arriva a parlare di esonizzazione degli elementi Alu quando, per un errore di splicing, parte della sequenza viene considerato una porzione esonica. Ci sono infatti circa 650 000 elementi Alu nelle regioni trascritte del genoma umano. Questi elementi contengono siti criptici di splicing, quindi costituiscono un costante pericolo per un'incorporazione aberrante all'interno dei trascritti maturi. Le sequenze Alu inserite sono antisenso (UUUUUUU-Right monomere-UUGUUU-Left monomere).

A sinistra troviamo una normale situazione di assenza di parte della sequenza Alu dall'mRNA maturo, a destra osserviamo un meccanismo di inclusione di Alu, non prevista, in seguito a mutazioni che hanno colpito la sequenza Alu stessa, generando siti di splicing criptici, nel trascritto maturo.



Nonostante rappresentino una seria minaccia all'integrità del trascrittoma, si possiedono poche conoscenze riguardo ai meccanismi molecolari che ne impediscono l'inclusione. Malattie come la cataratta congenita, il dismorfismo facciale e la sindrome della neuropatia sono causate dall'inclusione di elementi Alu intronici che interrompono drasticamente la struttura del trascritto, inficiando così sulla funzione della proteina risultante (Varon et al. 2003 "Nature Genetics"). Le mutazioni responsabili delle malattie appena elencate si verificano a carico del gene CTDP1 (CTD fosfatasi subunità 1, coinvolta nelle fasi iniziali della trascrizione).



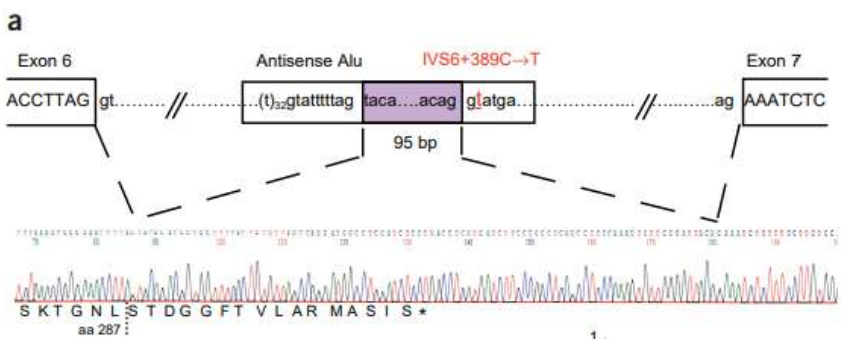


Figure 4 The Alu C→T mutation causes aberrant splicing of *CTDP1*. (**a**) The C→T substitution creates a donor splice site, which complies fully with the consensus sequence. In the resulting abnormal mRNA processing, the polypyrimidine tract of the antisense Alu element and an adjacent AG dinucleotide are recognized as an acceptor site (with splicing of the upstream part of intron 6), and the sequence surrounding the mutation serves as the next donor splice site. The insertion of 95 bp of the Alu sequence in the processed *CTDP1* transcript generates a premature termination signal 17 codons downstream of exon 6. (**b**) RT–PCR evidence of

Nonostante alcuni studi avessero precedentemente suggerito il coinvolgimento di fattori che agissero in trans, i meccanismi attraverso i quali le cellule si difendono dall'esonizzazione spuria degli elementi Alu sono rimasti sconosciuti fino ad oggi (2012).

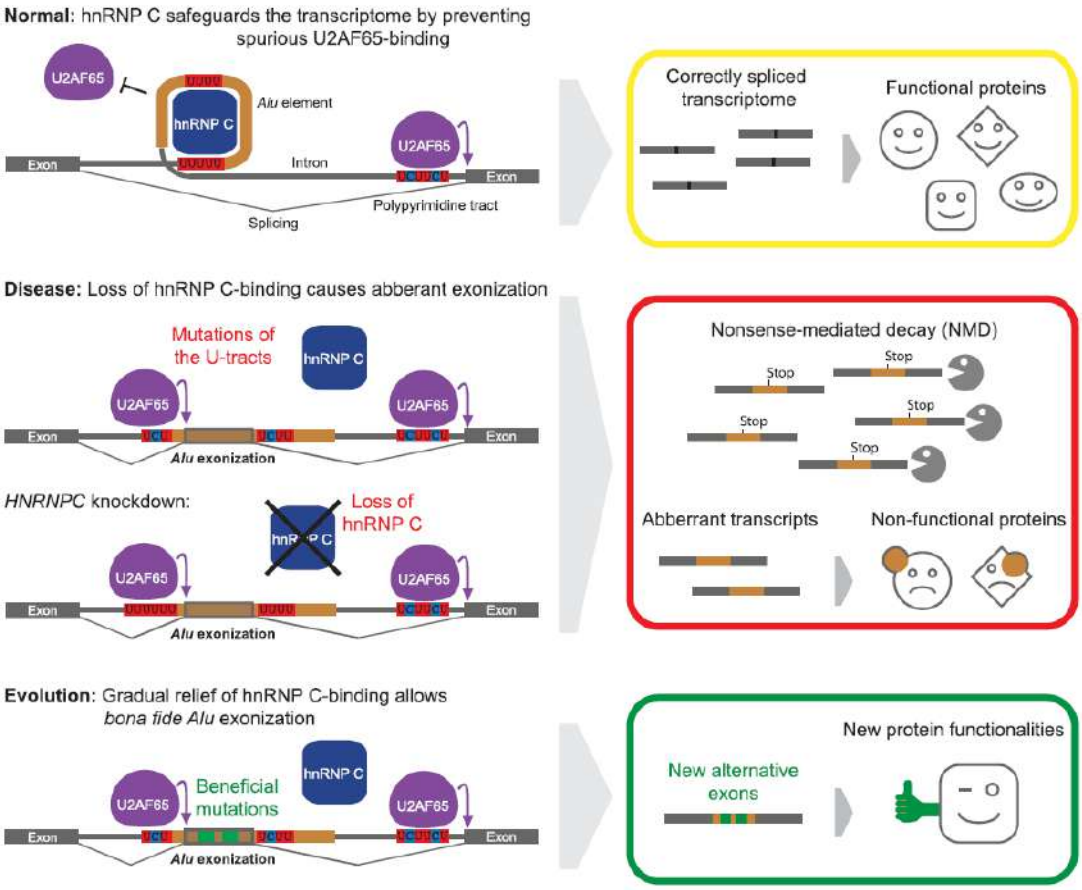


Figure 7. hnRNP C Safeguards the Transcriptome from the Exonization of *Alu* Elements
In normal cells, hnRNP C prevents recognition of the *Alu* elements through U2AF65, thereby ensuring accurate splicing. In the *HNRNPC* knockdown, U2AF65 can bind to the U-tracts and promote *Alu* exonization. Similarly, disease-associated mutations in the U-tracts can favor *Alu* exonization in the presence of hnRNP C by impairing hnRNP-C-binding. The resulting nonfunction transcripts are likely either targeted by nonsense-mediated decay (NMD) or give rise to nonfunctional proteins. Once an exon acquires beneficial changes during evolution, similar mutations accumulate to relieve hnRNP C repression, opening the possibility for new protein functionalities.

Il complesso hnRNP C previene l'esonizzazione degli elementi Alu mediante un legame forte con i loro tratti ricchi di uracile. I nostri esperimenti con minigeni, appunto, suggeriscono che la competizione di hnRNP C con U2AF65 nel legarsi ai tratti ricchi di U collocati a monte, nella regione 3' dell'Alu rappresenti un importante meccanismo per la repressione di Alu come esone.

Nelle cellule normali hnRNP C (proteina ribonucleare eterogenea nucleare C, che si lega all'RNA= impedisce il riconoscimento degli elementi Alu da parte di U2AF65 (U2 fattore ausiliare con peso molecolare di 65 kDa), assicurando così uno splicing corretto. Se la proteina hnRNP C viene "spenta", U2AF65 può legarsi ai tratti ricchi di U e favorire l'esonizzazione degli Alu. In maniera simile, mutazioni associate a malattie che abbiano luogo nei tratti ricchi di U possono promuovere l'esonizzazione di Alu in presenza di hnRNP C, attraverso un'incapacità di legame tra hnRNP C e Alu. I risultanti trascritti, non funzionali, possono andare incontro a una degradazione mediata da una mutazione non senso oppure producono proteine non funzionali. Una volta che un esone acquisisce mutazioni benefiche durante l'evoluzione, si accumulano mutazioni simili alla prima che portano a rilasciare la repressione operata da hnRNP C, garantendo la possibilità di nuove funzionalità per la proteina (si parla di esonizzazione di Alu bona fide). In conclusione, viene suggerito che hnRNP C svolga un ruolo chiave nel mantenere una condizione di salute nell'uomo, salvaguardando l'integrità del trascrittoma dagli effetti deleteri di un'esonizzazione spuria. Il ruolo potenziale di hnRNP C come guardiano competitivo del trascrittoma, come altri tipi di elementi trasponibili rimane ancora da determinare.

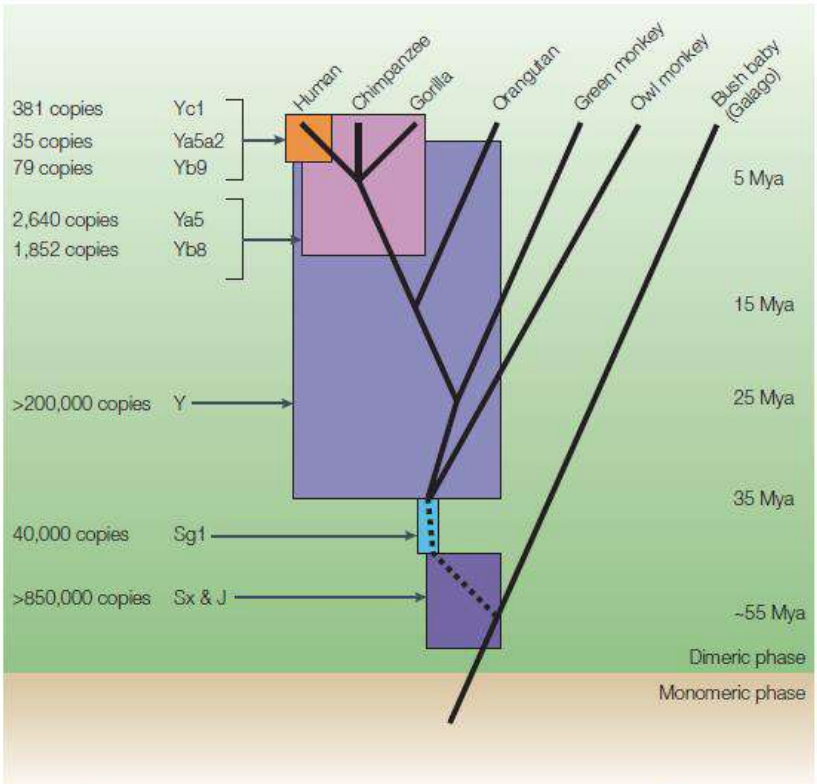


Figure 2 | **The expansion of Alu elements in primates.** The expansion of Alu subfamilies (Yc1, Ya5a2, Yb9, Yb8, Y, Sg1, Sx and J) is superimposed on a tree of primate evolution. The expansion of the various Alu subfamilies is colour coded to denote the times of peak amplification. The approximate copy numbers of each Alu subfamily are also noted. Mya, million years ago.

E' avvenuta un'espansione degli elementi Alu all'interno dell'albero evolutivo dei Primati. Gli Alu, inizialmente monomerici, sono divenuti dimerici. Al contempo, si è ridotto il numero di copie presente nel genoma dei vari organismi.

Nel 2014 sono stati identificati anche siti target di triplicazione (TSTs), dove due elementi Alu in tandem condividono una ripetizione diretta. Perciò sono presenti tre copie della ripetizione. Ogni elemento Alu, infatti, è fiancheggiato da due porzioni identiche di DNA, le TSDs. L'identità di sequenza tra i due segmenti è causa di crossing-over ineguale tra i due cromosomi omologhi, provocando la duplicazione di un elemento Alu e alla formazione di un sito target di triplicazione nel genoma umano. La ricombinazione ineguale tra due sequenze Alu (ricombinazione Alu-Alu) produce una quadruplicazione. Altri multipli della TSD sono spiegati da questo meccanismo.

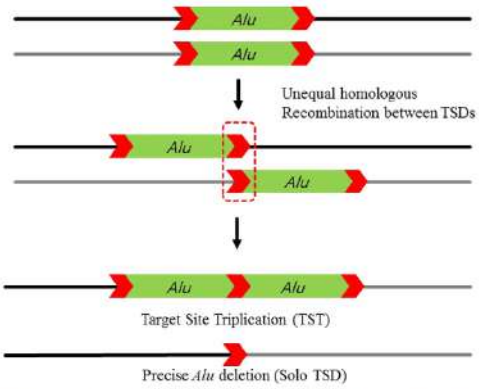


Fig. 1. The mechanism of target site triplication. Each Alu elements are flanked by two identical DNA segments, TSDs. The sequence identity between the two segments causes unequal homologous recombination between them, leading to the duplication of an Alu element and target site triplication (TST) in the human genome.

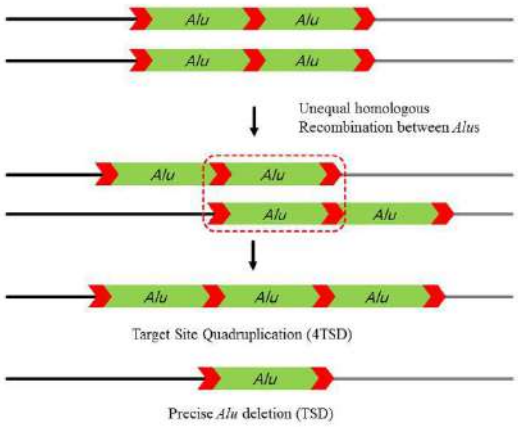
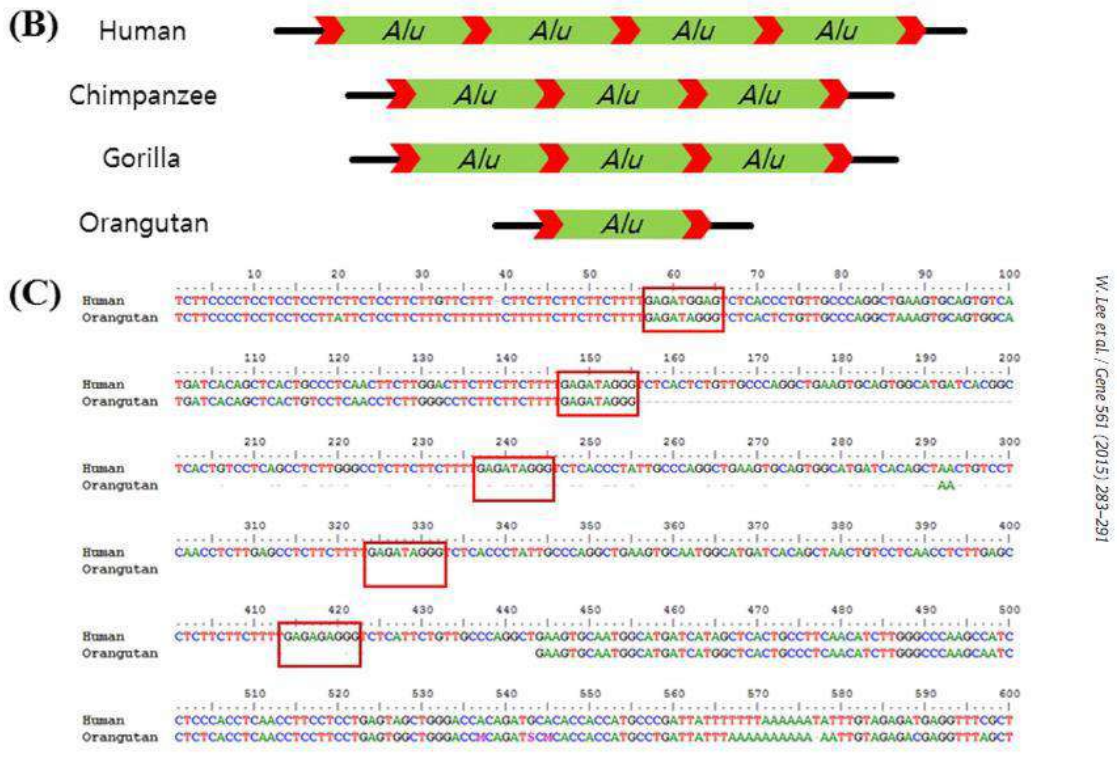


Fig. 3. The mechanism of quadruplication. Unequal homologous recombination between two Alu sequences (Alu–Alu recombination) creates quadruplication. Other multiple TSD was also explained by this mechanism.

Nel genoma umano ci sono circa il doppio di elementi Alu che nel genoma dello scimpanzè. Ciò significa che forse nello scimpanzè questi elementi sono stati meno attivi.

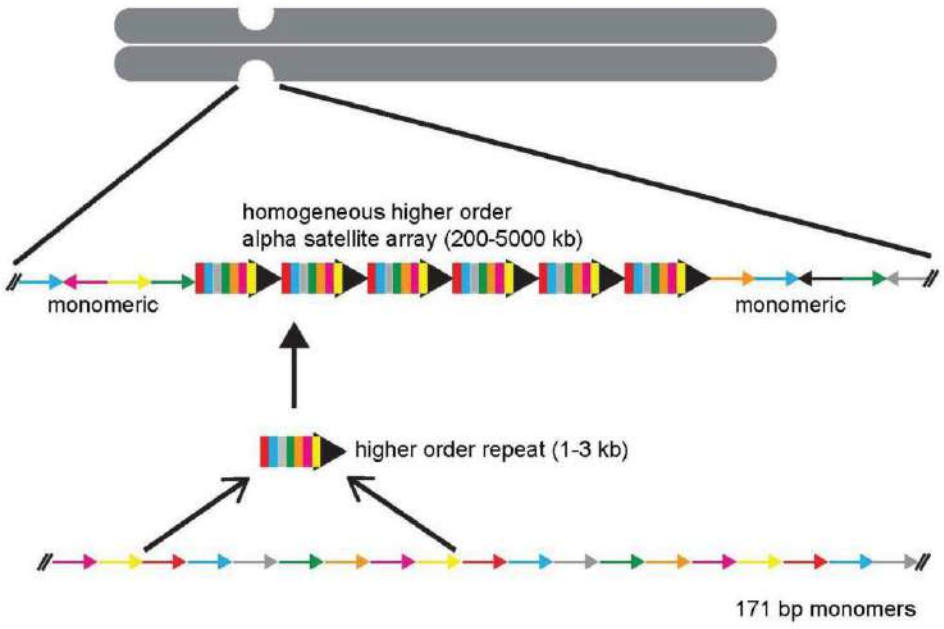


L'elemento Alu è stato un'importante fonte di nuovi esoni durante l'evoluzione dei Primati. Più del 25% degli esoni analizzati con il sequenziamento dell'RNA hanno stimato livelli di inclusione nel trascritto pari almeno al 50% nel cervelletto umano, mostrando una globale affermazione degli esoni Alu nei geni umani. Geni codificanti per fattori di trascrizioni con dominio a dita di zinco possiedono livelli significativamente più alti di esonizzazione di Alu. Un'importante conseguenza funzionale dell'esonizzazione di Alu è l'evoluzione specifica di una discendenza nella regolazione traduzionale. Inoltre, la formazione preferenziale e l'affermazione degli elementi Alu nei geni per le "dita di zinco" suggeriscono che l'esonizzazione di Alu potrebbe aver influito globalmente sull'evoluzione dei Primati e dei trascrittomi umani attraverso la regolazione di prodotti proteici dei principali regolatori della trascrizione in specifiche discendenze.

Genetica forense

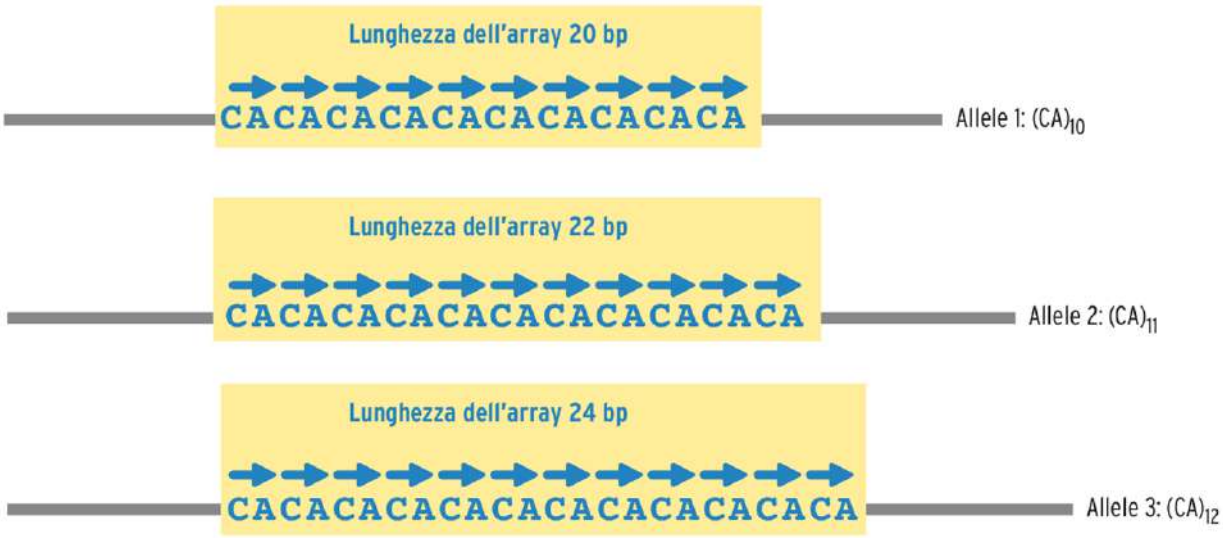
La genetica forense ha lo scopo di identificare colpevoli di crimini, risolvere casi di paternità incerta, predire la popolazione di origine di un DNA (es. Africano, Asiatico, ecc.), dare un'identità a cadaveri ad es. vittime di catastrofi naturali, disastri aerei e simili eventi. La genetica forense si basa sulla presenza di diversità genetica tra individui, tra popolazioni, tra specie. Nella maggior parte delle applicazioni la genetica forense sfrutta la diversità genetica tra individui. Agisce a livello di diverse porzioni del genoma: sugli autosomi, sui cromosomi X e Y, sul DNA mitocondriale. I marcatori più utilizzati in ambito forense sono localizzati sugli autosomi. Vengono considerati differenti tipi di variazioni: i microsatelliti (short tandem repeats, o STS), i polimorfismi a singolo nucleotide (SNPs), i minisatelliti (VNTR), l'inserzione o delezione di basi (indel), le varianti nel numero di copie (CNVs) e i polimorfismi "citogenetici". I marcatori più utilizzati in ambito forense sono i microsatelliti.

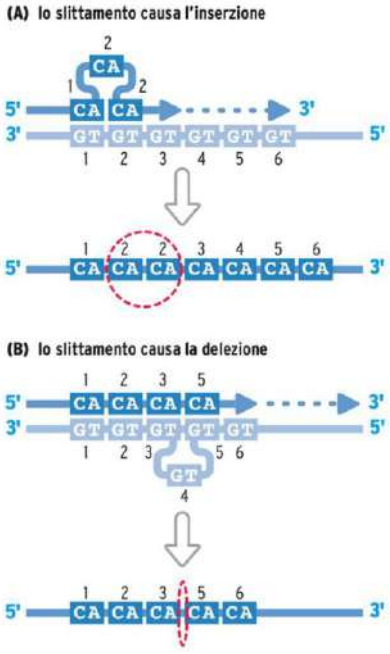
Le ripetizioni in tandem nel DNA vengono classificate in tre categorie, in base alle dimensioni delle porzioni che contengono tali ripetizioni. Si hanno:



- DNA satelliti, con un array (lunghezza) da 20 a centinaia di kb. Si osservano nelle regioni eterocromatiniche e in prossimità dei centromeri;
- DNA minisatelliti, con un array da 100 bp a 20 kb. Si osservano nei telomeri e nelle regioni subtelomeriche;
- DNA microsatelliti, con un array di meno di 100 bp. Sono uniformemente distribuiti nel genoma.

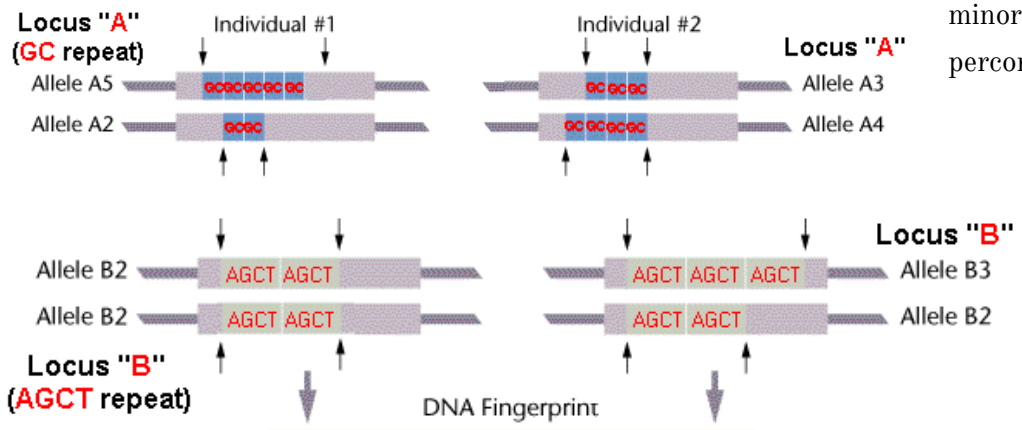
- I microsatelliti, in particolare, sono i marcatori genetici più utilizzanti per tre ragioni: sono altamente polimorfici, sono abbondanti nel genoma, sono uniformemente distribuiti.

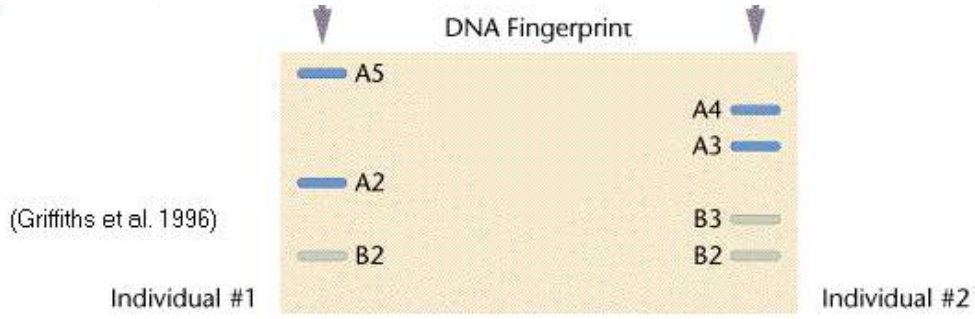




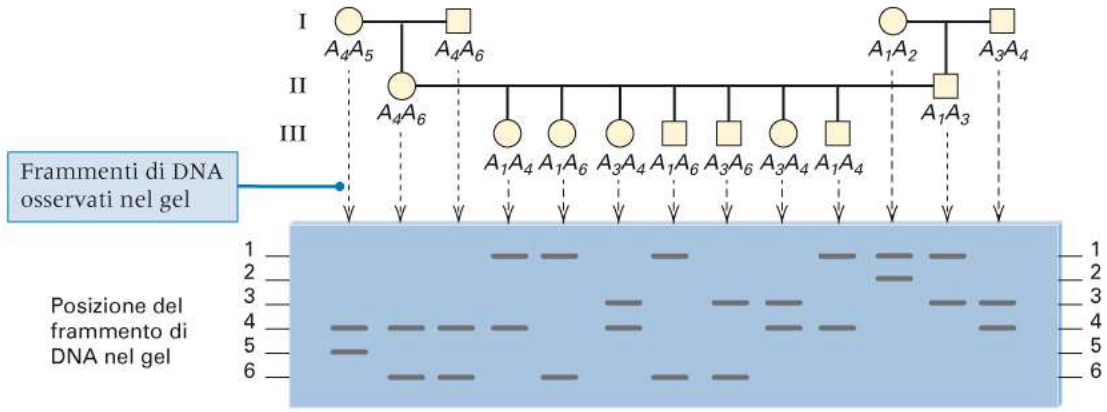
La variazione nel numero di ripetizioni, diverso per ogni individuo, viene causata da uno slittamento dell'RNA polimerasi nel processo di trascrizione. La polimerasi può, erroneamente, trascrivere due volte la stessa ripetizione, incrementando di uno il numero di repeat, ma può anche escludere una delle ripetizioni, diminuendone il numero totale.

Di seguito viene riportato il metodo con il quale si analizzano i microsatelliti presenti in due individui diversi, l'individuo #1 e l'individuo #2, per due loci diversi (A e B). Nel locus A si contano le ripetizioni della coppia GC, mentre nel locus B quelle del quartetto AGCT. La sigla composta da numero più lettera indica per ciascun allele di quel locus quante ripetizioni si presentano. La sigla "A5", ad esempio, significa che l'individuo presenta 5 ripetizioni di GC. I frammenti allelici, dopo essere stati amplificati e digeriti, vengono fatti correre su un gel di agarosio. Maggiore è il numero di ripetizioni contenute, minore sarà il fronte percorso dal segmento.





Due individui, fatta eccezione per i gemelli monozigoti, non possono avere la stessa sequenza di basi nel loro DNA. Per questa ragione sono molto importanti, nella genetica forense, le regioni polimorfiche del DNA: i profili allelici possono essere individuo-specifici, consentendo la determinazione dell'identità genetica. L'analisi simultanea di numerosi microsatelliti comporta milioni di differenti combinazioni o profili di DNA e aumenta la possibilità di distinguere individui diversi.



Se si osserva una sola banda per un locus, o l'individuo presenta lo stesso allele su entrambi i cromosomi, oppure l'amplificazione non è avvenuta correttamente e bisogna ricominciare da capo. Questi sistemi servono per l'identificazione di un colpevole in occasione di un crimine.

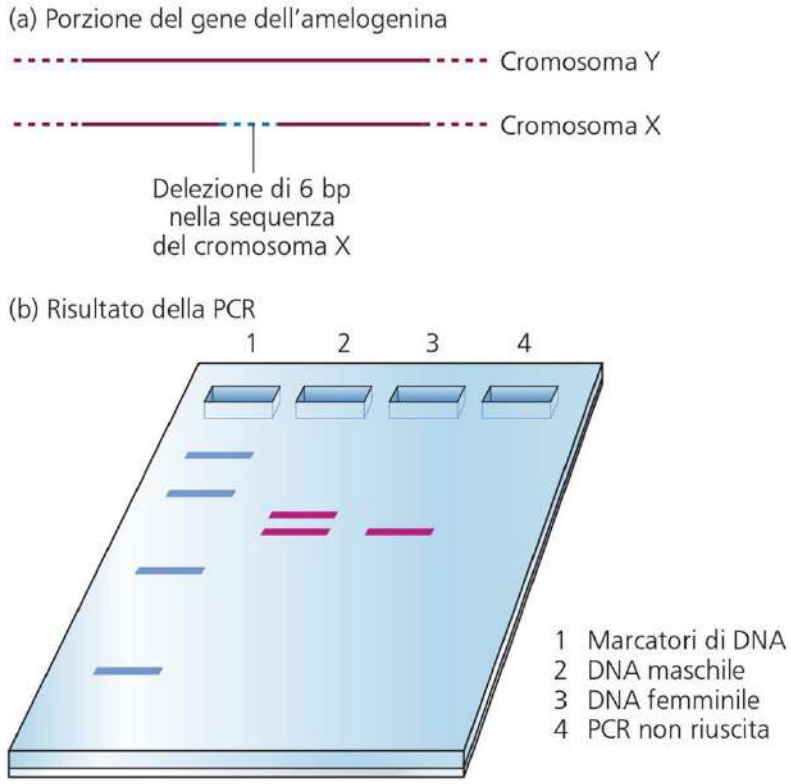
TABELLA 16.2

Le brevi ripetizioni in tandem (STR) usate per la tipizzazione forense del DNA dal Federal Bureau of Investigation (FBI; Stati Uniti), dal Forensic Science Service (FSS; Regno Unito) e dall'Interpol (Europa)

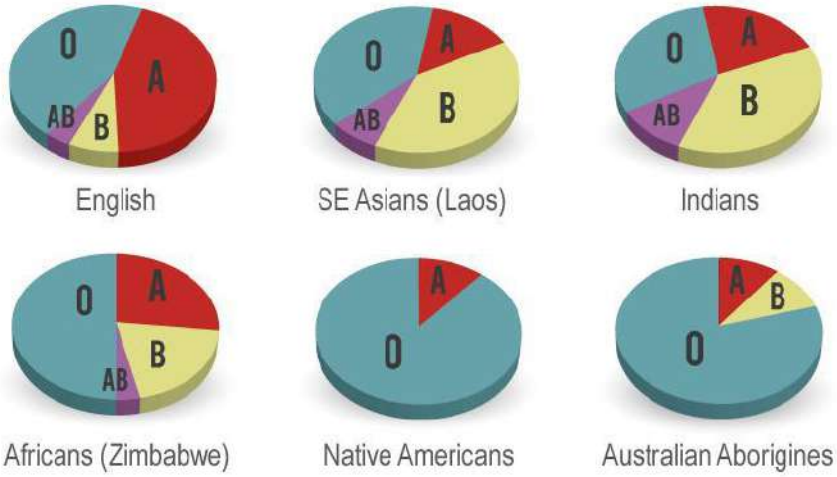
	Ripetizione	Numero di alleli	FBI	FSS	Interpol
D2S1338	[TGCC][TTCC]	20		✓	
D3S1358	[TCTG][TCTA]	10	✓	✓	✓
D5S818	AGAT	10	✓		
D7S820	GATA	11	✓		
D8S1179	[TCTA][TCTG]	10	✓	✓	✓
D13S317	TATC	8	✓		
D16S539	GATA	8	✓	✓	
D18S51	AGAA	15	✓	✓	✓
D19S433	AAGG	19		✓	
D21S11	[TCTA][TCTG]	69	✓	✓	✓
CSF1P0	TAGA	15	✓		
FGA	CTTT	19	✓	✓	✓
THO1	TCAT	7	✓	✓	✓
TPOX	GAAT	7	✓		
vWA	[TCTG][TCTA]	10	✓	✓	✓
Amelogenina	106 bp/112 bp	2	✓	✓	✓

Le parentesi quadre indicano i microsatelliti con profili complessi di ripetizione.

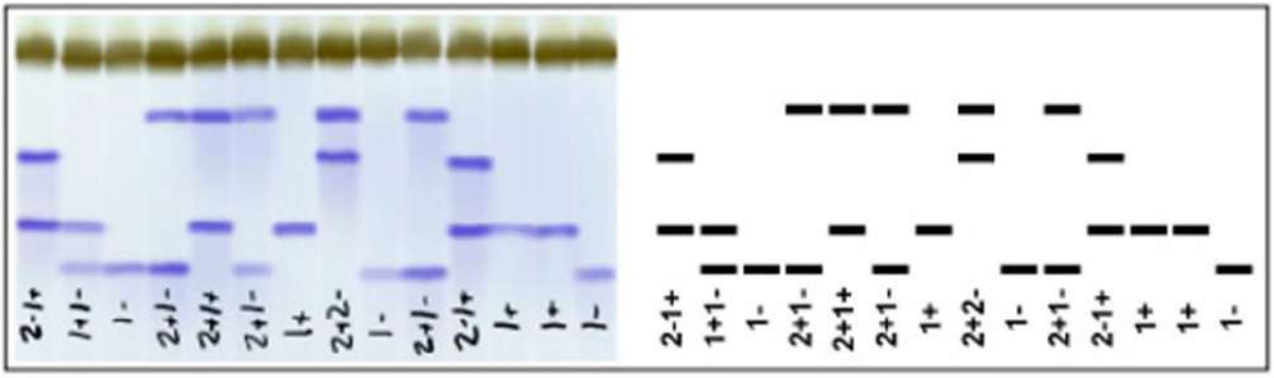
Il gene dell'amelogenina è usato per determinare il sesso. Il gene si trova su una regione omologa sui cromosomi X e Y e quindi è presente in due copie sia nei maschi che nelle femmine. La lunghezza del prodotto di PCR ottenuto dal cromosoma Y è di 112 bp e la lunghezza di quello del cromosoma X è di 106 bp.



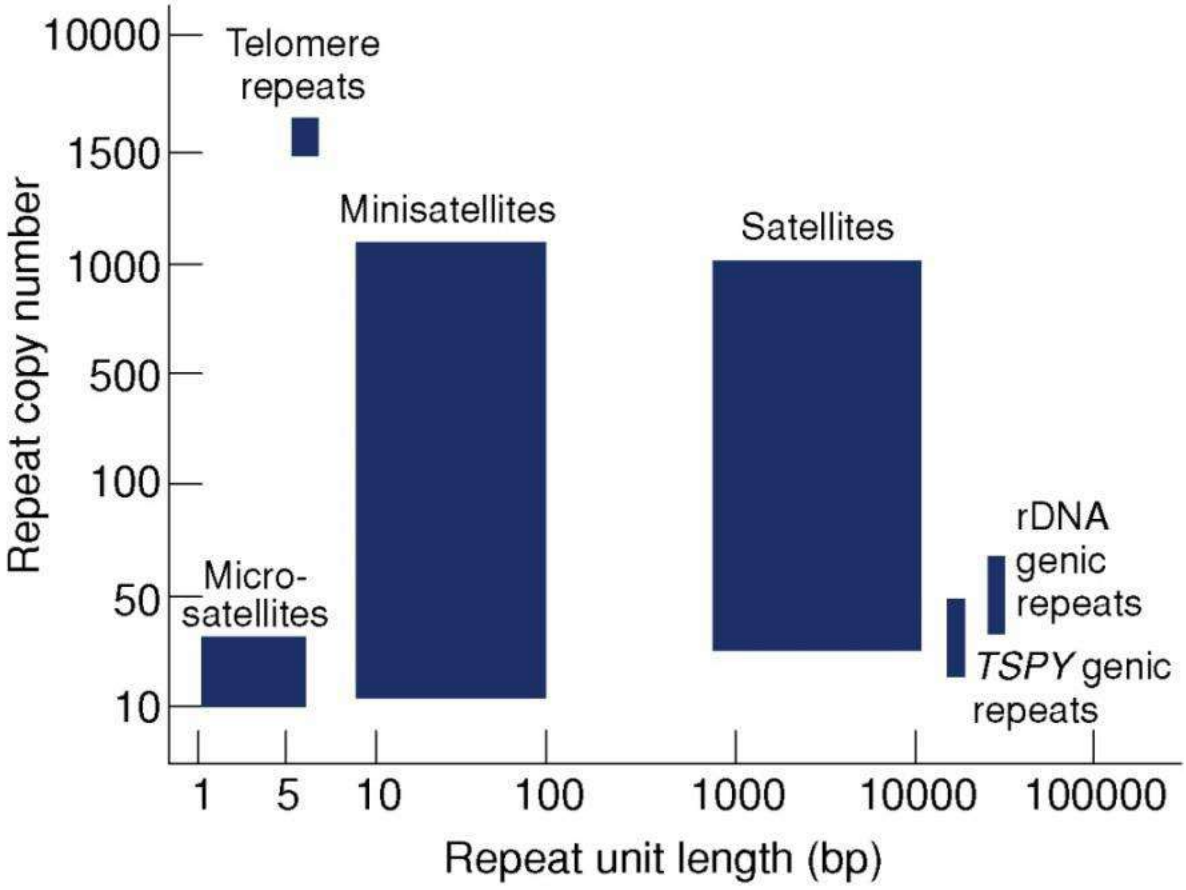
Il primo sistema polimorfico utilizzato in ambito forense, dove sussisteva un sospetto di omicidio, fu quello che considerava il sistema di gruppo sanguigno AB0. Venne ideato nel 1915 dal Professore Leone Lattes, docente di Medicina Forense presso l'Università di Torino, che scoprì un metodo per verificare il gruppo sanguigno (A, B, AB, 0) a partire da una macchia di sangue asciutto. Questo sistema contemplava dunque 4 fenotipi, comportava una discreta rapidità di esecuzione, ma permetteva soltanto possibilità di esclusione e, inoltre, aveva scarso potere discriminatorio.

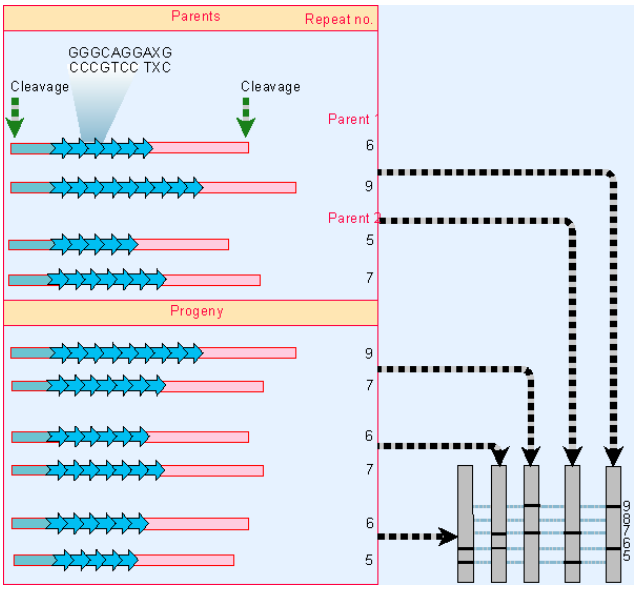


Dagli anni ’60 agli anni ’80 venne seguito un procedimento che prendeva in considerazione i polimorfismi proteici o enzimatici: differenti forme dello stesso enzima (isoenzimi) venivano evidenziate mediante elettroforesi su agar o acrilammide, sfruttando la differente carica degli isoenzimi per la separazione. Gli alleli per ciascun enzima erano però in numero ridotto (da 2 a 4). La procedura richiedeva tempi di esecuzioni lunghi. Aveva in effetti un maggiore potere di discriminazione (quando più isoenzimi venivano combinati) rispetto al sistema AB0, ma i livelli di proteine contenuti nei reperti analizzati erano spesso bassi e mostravano spesso instabilità.

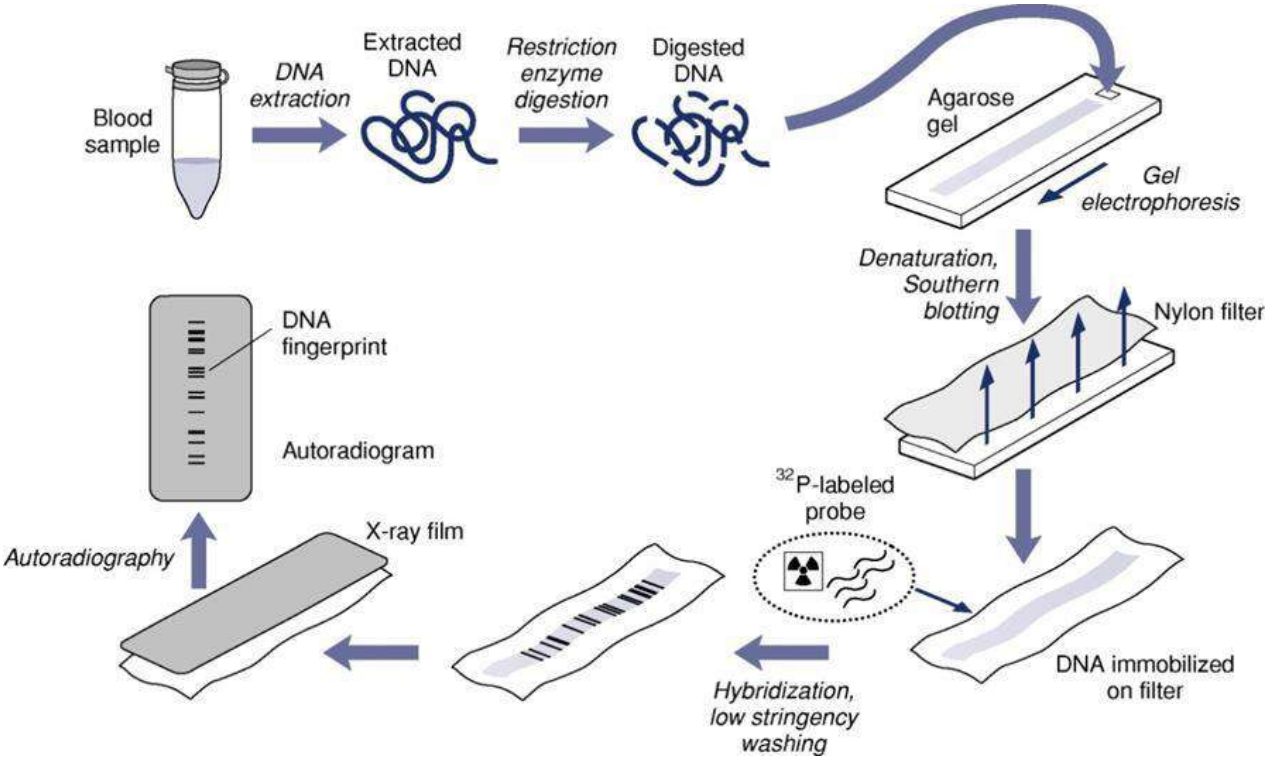


Il 1985 fu l’anno dell’avvento dei minisatelliti con l’idea del “DNA fingerprinting” da parte di Alec Jeffreys.



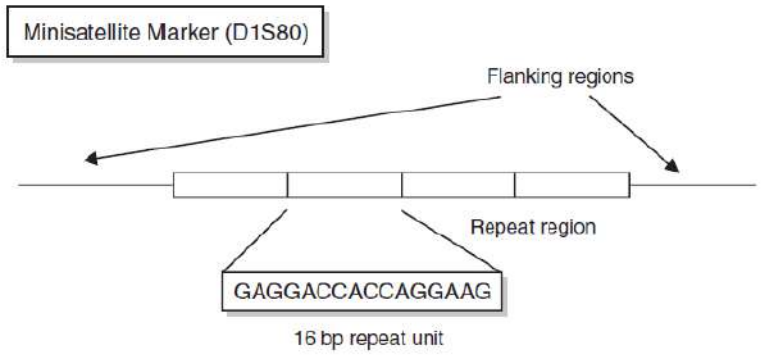


Da un campione di sangue si estrae il DNA, lo si fa tagliare da un enzima di restrizione, che lo digerisce. Si fanno correre i frammenti ottenuti su una piastra per l'elettroforesi su gel di agarosio. Avviene la denaturazione, dopodichè si esegue il Southern blotting utilizzando un filtro in nylon. Una volta immobilizzato il DNA sul filtro, ci si serve di una sonda marcata con P32 radioattivo, che riconosce un motivo "core" comune al minisatellite. Si lascia che abbia luogo l'ibridazione sonda-DNA e si effettua un lavaggio a bassa stringenza. Si imprime il filtro su una lastra a raggi X che viene esposta in un autoradiografia, producendo un autoradiogramma che permette di leggere i risultati ottenuti. Un innegabile vantaggio era l'elevato potere di discriminazione, ma questa tecnica presentava innumerevoli svantaggi: la sensibilità era limitata (servivano almeno 50 nanogrammi fino a 500 ng ammessi); il protocollo da seguire era "time-consuming" e non automatizzabile, riconosceva solo porzioni di DNA ad elevato peso molecolare (ciò rendeva la tecnica inadatta per campioni degradati) e considerava alleli spesso difficili da distinguere.



Nel 1983 e nel 1986 vennero compiuti due delitti con violenza nel Regno Unito. Si scoprì subito che il gruppo sanguigno dell'omicida era A. Nel 1986 un ragazzo venne incriminato per questi due fatti delittuosi. Fu richiesto a Jeffreys l'analisi del DNA fingerprinting, coinvolgendo 5000 ragazzi che si

sapeva avessero gruppo sanguigno A. Si eseguì dunque uno screening di massa, ma nessun DNA coincideva con il fingerprinting. Nel 1987, per caso, si seppe che un ragazzo del luogo aveva chiesto ad un amico di partecipare allo screening in sua vece. Il profilo del DNA del colpevole risultò identico a quello isolato sulla scena del crimine.

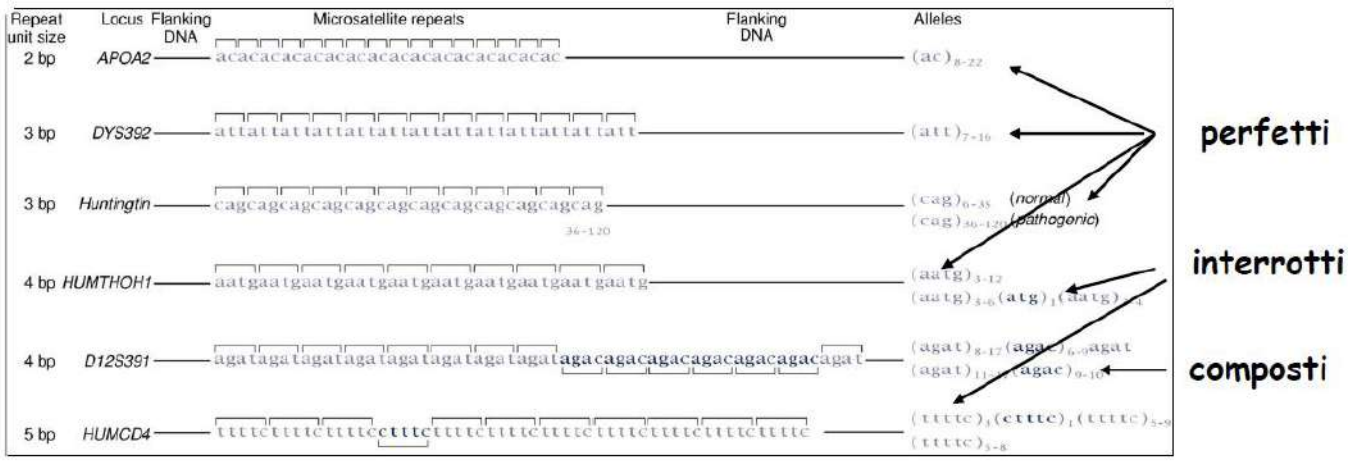


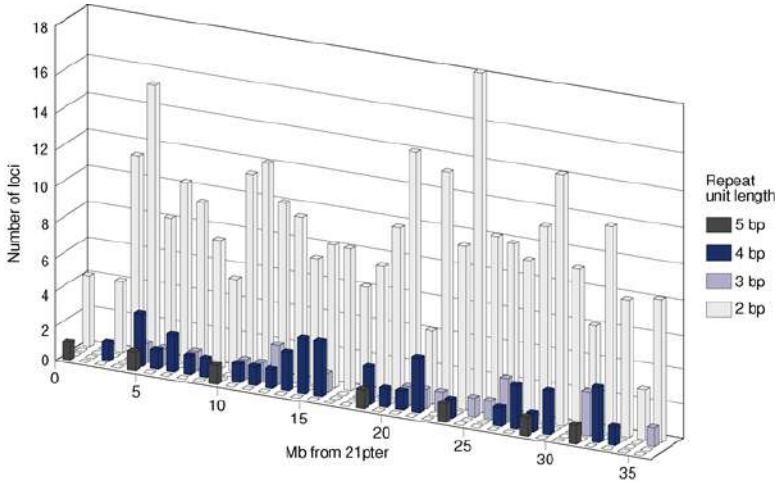
Nel 1994 venne individuato il minisatellite D1S80, amplificabile mediante PCR. Bastavano quantità minime di DNA e, inoltre, era uno dei marcatori multiallelici con elevato grado di eterozigosità. Il numero di repeats variava da 14 a 41. L'amplicone prodotto aveva una taglia compresa tra 400 e 800 bp. Il metodo era semplice, perché si componeva di PCR e elettroforesi, la PCR permetteva un'elevata sensibilità e si usufruiva di un potere di discriminazione elevato per un singolo locus. L'amplicone era, però, piuttosto grande, dunque si ponevano problemi nel caso in cui fosse stato estratto DNA degradato. Inoltre, vennero riscontrate difficoltà per l'analisi in multiplex e per l'automazione.

Dal 1995 si passò ai microsatelliti, che sono usati tuttoggi. Nei Primati essi rappresentano l'1-2% del genoma. Un esempio di microsatellite utilizzato in ambito genetico forense è D16S539, che amplifica una serie di repeat GATA.

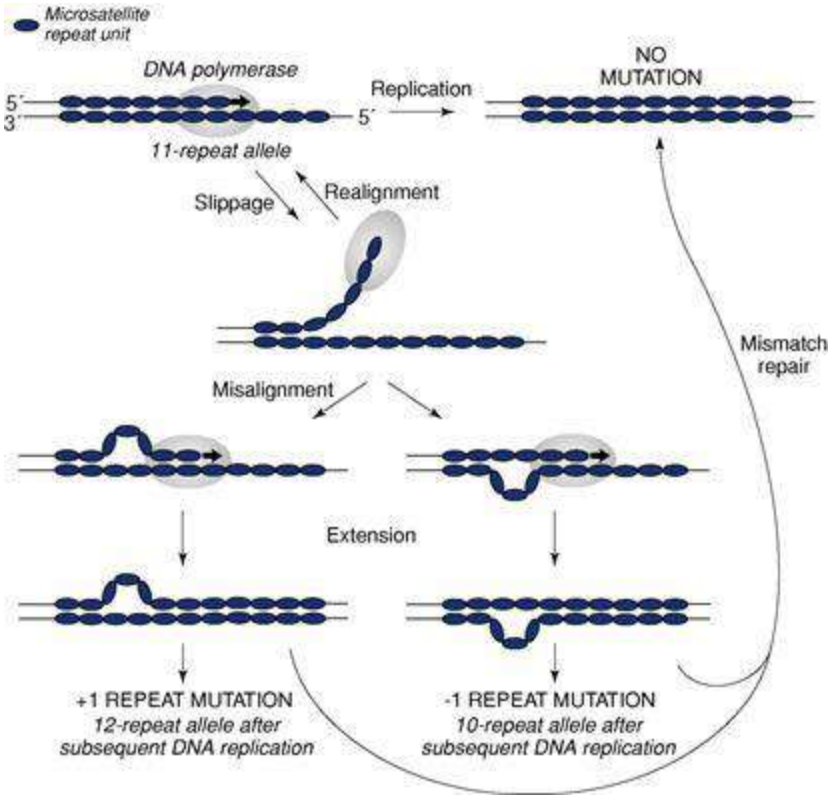
GGAGGATGACTGTGTTCCCACTCTCAGTCCTGCCGAGGTGCCTGACAGCCCTG
CACCCAGGAGCTG**GGGGGTCTAAGAGCTTGTAAAAAG**TGTACAAGTGCCAGAT
GCTCGTTGTGCACAAATCTAAATGCAGAAAAGCACTGAAAGAAGAATCCCGAA
AACCACAGTTCCCATTTTTATATGGGAGCAAACAAAGCAGATCCCAAGCTCTT
CCTCTTCCCTAGATCAATACAGACAGACAGACAGGTG**/gata/gata/gata/**
gata/gata/gata/gata/gata/gata/gata/gata/TCATTGAAAGACA
AAACAGAGATGGATGATA**GATACATGCTTACAGATGCACACACAAAC**GCTAAA
TGGTATAAAAATGGAATCACTCTGTAGGCTGTTTTACCACCTACTTTACTAAA
TTAATGAGTTATTGAGTATAATTTAATTTTATATACTAATTTGAAACTGTGTC
ATTAGGTTTTTAAGT

I microsatelliti possono essere perfetti, quando sono formati da ripetizioni continue di una certa serie di basi, interrotti, se le repeat sono inframezzate da alcune basi che separano il "tandem", composti, se si combinano ripetizioni contigue di diverse sequenze.

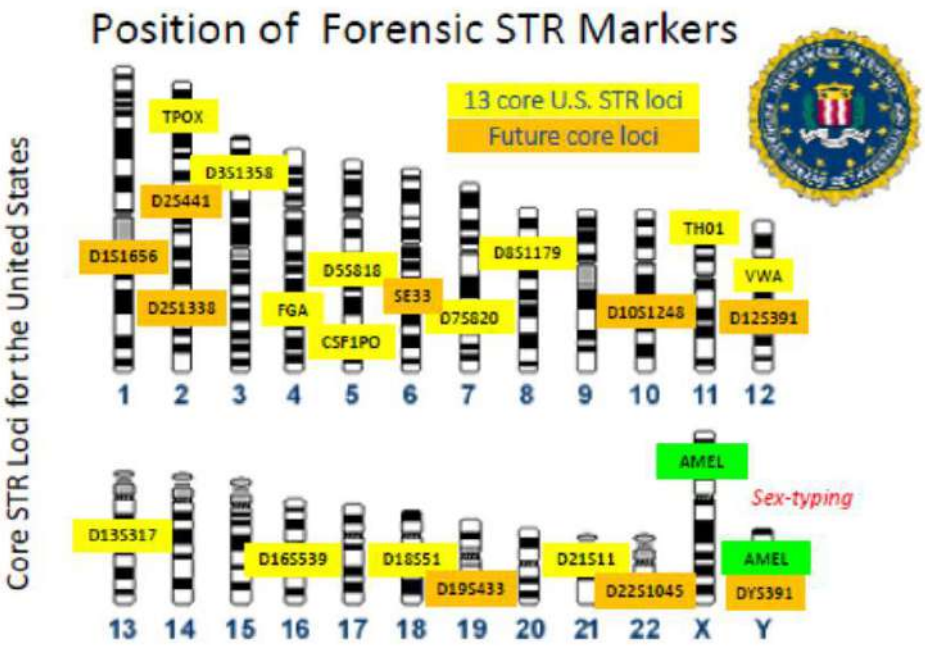




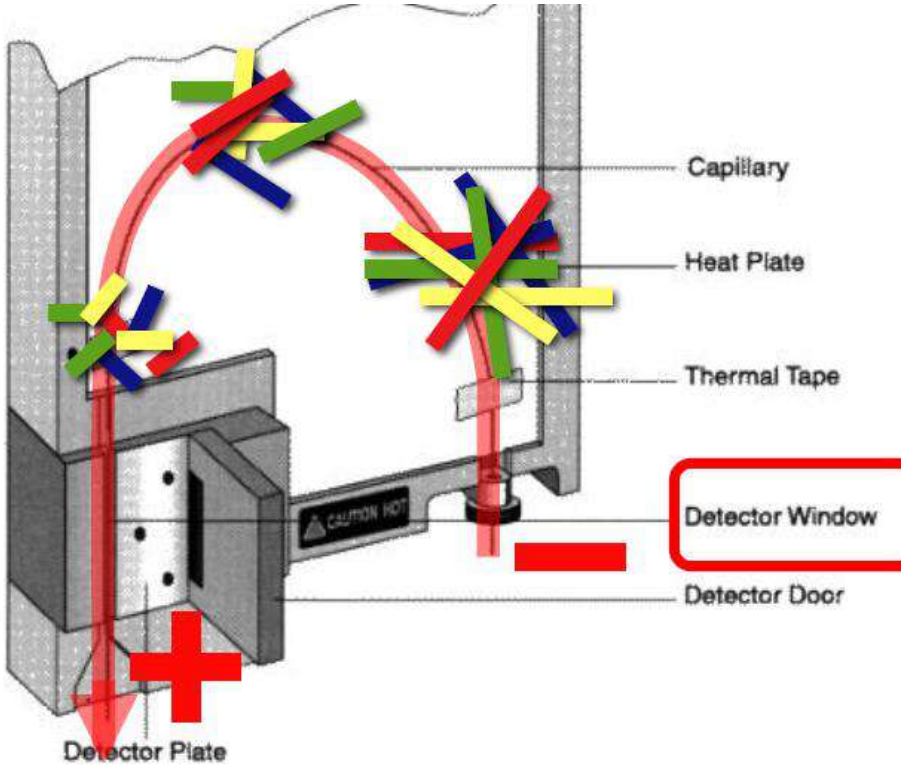
La variazione nel numero di repeat, dovuta a uno slittamento della polimerasi, può essere anche riparata alla luce del mismatch che si verifica in seguito allo slittamento.

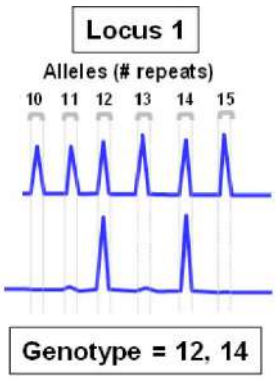


Il CODIS (Combined DNA Index System) è un programma dell'FBI in supporto alla giustizia che si basa sullo sviluppo e l'utilizzazione di databases di profili genetici. Esso ha disegnato un pannello di 13 microsatelliti "core". In UK e alcuni altri stati Europei si utilizza un core di 10 STR, 8 dei quali in comune con il CODIS. I 13 loci sono tutti «tetranucleotide repeats». L'analisi dei 13 loci del CODIS porta a profili genetici la cui "Random Match Probability" media è minore di 1 su un milione di milioni (10^-12).



Per generare un profilo STR si procede all'estrazione e quantificazione del DNA del campione, che può essere costituito da sangue, liquido seminale, capelli, saliva, urine. Si passa all'amplificazione mediante PCR multiplex (con formazione di un complesso primer-fluorocromo). Ciò comporta un possibile rischio di contaminazione da parte dell'operatore, quindi esiste un database dove sono schedate le informazioni genetiche del personale di laboratorio. Il prodotto di PCR fluorescente viene denaturato.

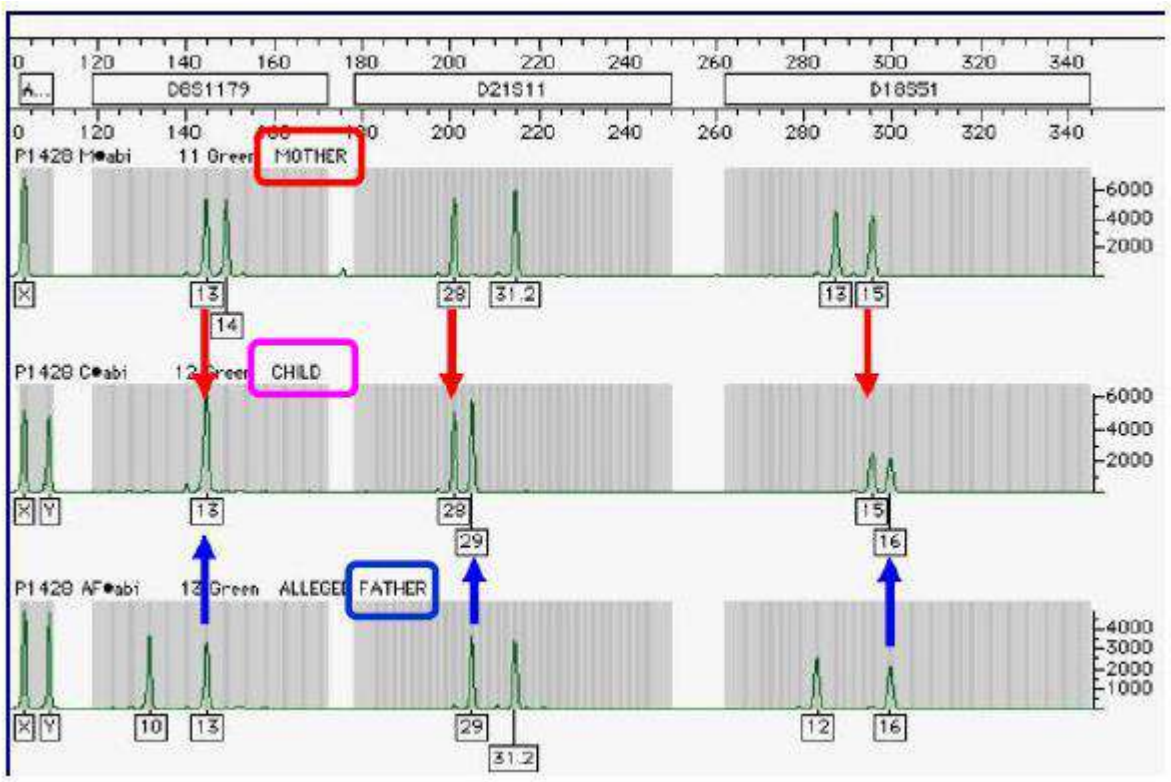


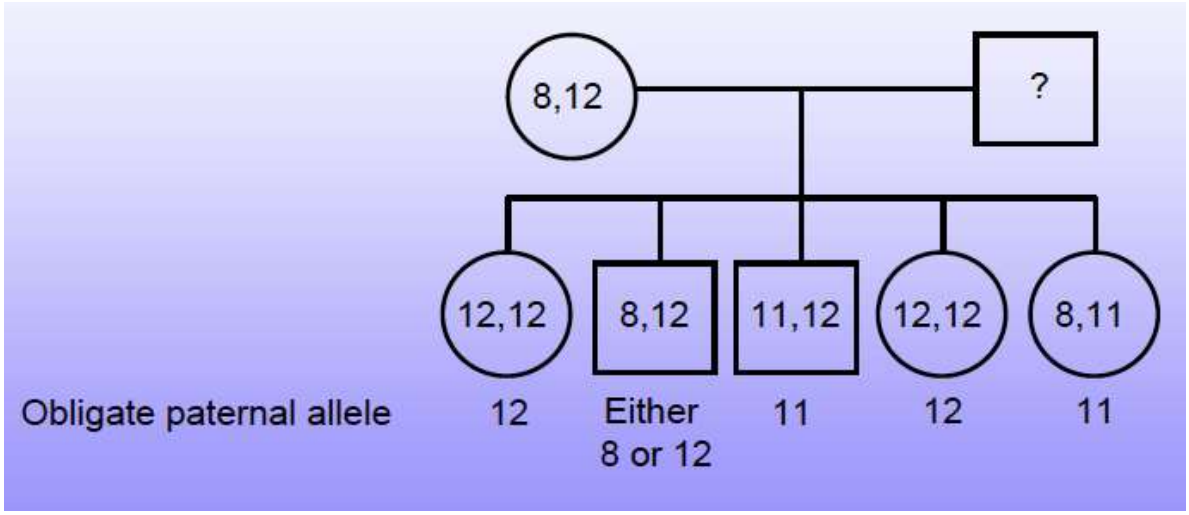


All'interno dell'apparecchio per l'elettroforesi capillare viene applicato un voltaggio elettrico e il prodotto amplificato viene elettroiniettato nel capillare. Il DNA migra nel polimero contenuto nel capillare. Le molecole si separano in base alla loro lunghezza. Quando il frammento arriva in corrispondenza di una finestra nel capillare, una sorgente laser eccita il fluorocromo legato al prodotto di PCR.

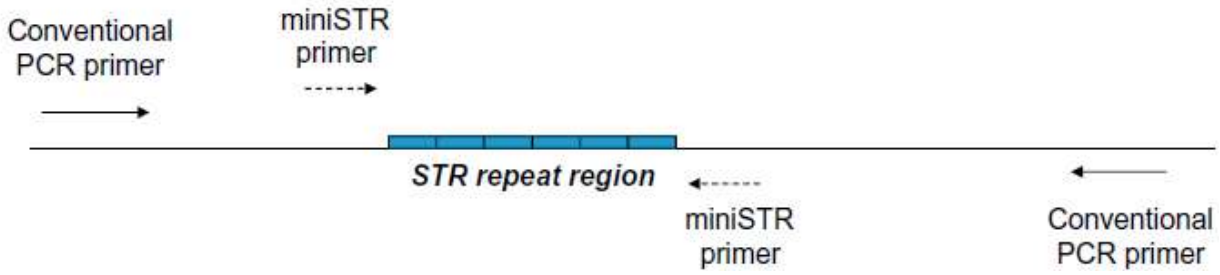
Il genotipo di un campione si ottiene mediante confronto del profilo con quello di un ladder allelico.

Nei test di paternità sono possibili due tipi di risultato: uno di compatibilità, ovvero tutti gli "alleli obbligati paterni" del figlio sono presenti nel padre putativo, oppure di esclusione, se più "alleli obbligati paterni" del figlio non sono presenti nel genotipo del padre putativo.





Se il campione del DNA degradato sono diverse le procedure da adottare: si può amplificare una SNP, i cui ampliconi sono lunghi meno di 100 basi; si può analizzare il DNA mitocondriale, il cui elevato numero di copie rispetto al nucleare aumenta la probabilità che i primer si appaino ad un segmento di DNA "intatto"; si ricorre a "mini-STRs", cioè STR con primer ridisegnati in modo da ottenere ampliconi più piccoli di quelli utilizzati normalmente.



Se il DNA è scarso (minore di 100-200 pg, oppure il DNA di 15-30 cellule diploidi), si analizza il DNA mitocondriale.

I database di profili sono essenzialmente collezioni "informatiche" di singoli profili (genotipici) relativi a più marcatori (es. profili di microsatelliti Y, profili di microsatelliti autosomici, sequenze di mtDNA). L'importanza dei database in genetica forense è aumentata dagli anni '90 ad oggi e la loro tipologia si è diversificata nel tempo. Esistono:

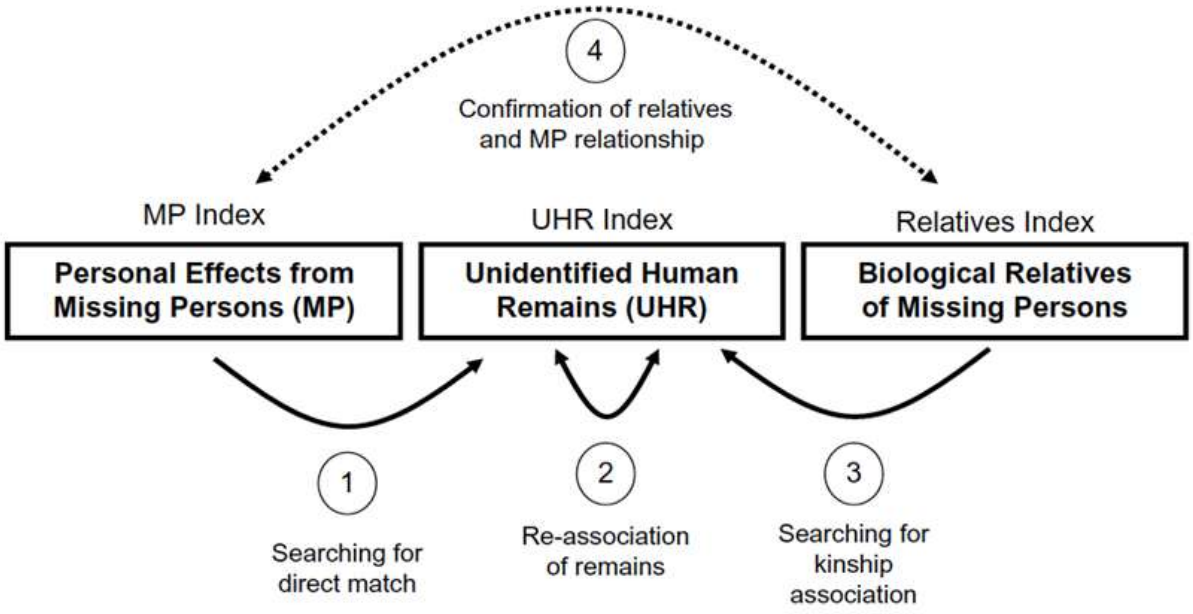
-database di profili di «offenders» (rei, imputati, indagati....);

-database di profili di reperti forensi (da DNA lasciato sulla scena del reato);

-database di profili di DNA di persone scomparse (DNA dei loro effetti personali);

-database di profili di DNA di parenti di persone scomparse (MP);

-database di profili di DNA di resti umani non identificati (UHR).



Nel processo di identificazione delle persone disperse al World Trade Center in seguito all'attacco terroristico dell'11 settembre 2001 sono stati generati 52 000 profili STR, 44 000 profili di mtDNA e 17 000 profili SNPs. 850 delle 1594 vittime sono state identificate grazie all'analisi del DNA.

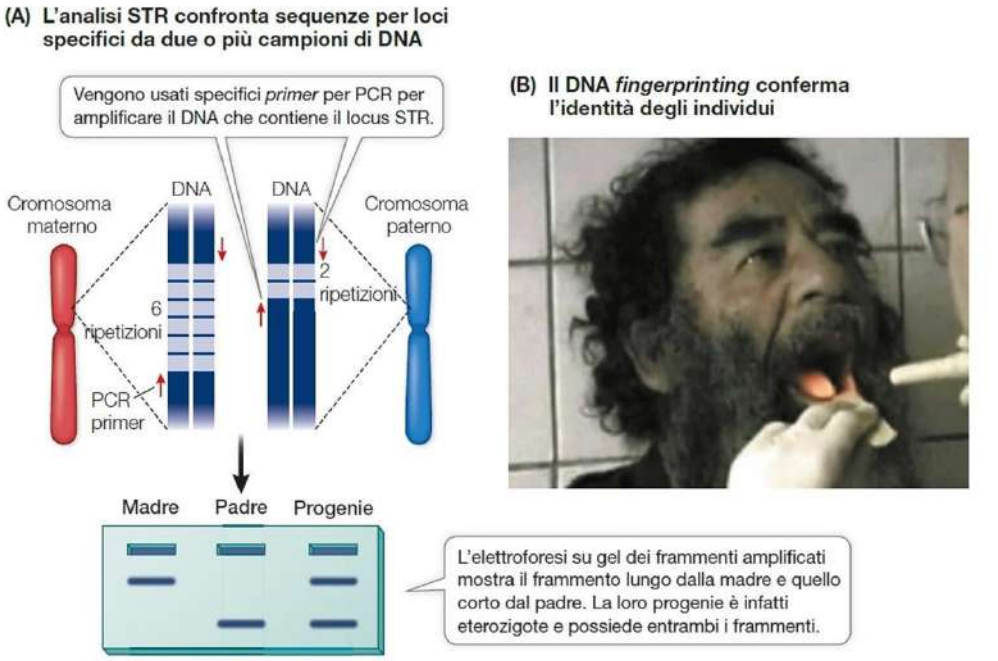


Figura 15.13 DNA *fingerprinting* con microsatelliti (STR)
(A) Un particolare locus STR può essere analizzato per determinare il numero di sequenze ripetute ereditate da un individuo da ogni genitore. I due alleli possono essere identificati su un gel tramite elettroforesi sulla base della loro dimensione. Quando si analizzano contemporaneamente parecchi loci STR, il pattern osservato può permettere l'identificazione totalmente certa di un individuo. (B) Quando il dittatore Saddam Hussein fu catturato in Iraq, fu raccolto nell'interno della guancia un campione del suo epitelio per fare il DNA *fingerprinting*. Un confronto con i DNA *fingerprinting* dei parenti fornì agli scienziati militari la prova che l'uomo in questione era proprio Saddam Hussein.

? Dovremmo raccogliere il DNA di tutti già alla nascita per assegnare una «carta d'identità genetica» da utilizzarsi negli screening di malattie genetiche? Quali sarebbero vantaggi e svantaggi di un'operazione del genere?

Il caso dell'omicidio di Yara Gambirasio (26/11/2010)

Il cadavere venne ritrovato il 26 febbraio 2011 e riportava numerose ferite da taglio. Il 9 maggio venne scoperta una traccia ematica, dalla quale venne isolato un profilo genetico maschile di ottima qualità (cromosoma Y + loci microsatellitari), di un individuo che venne chiamato "ignoto 1". Si procedette quindi alla ricerca di compatibilità dell'aplotipo del cromosoma Y (con 16 microsatelliti presi in esame) con i soggetti maschili appartenenti alla famiglia e dei conoscenti della vittima. Si confrontano i dati ottenuti con l'elenco dei nomi di chi, nelle ore della scomparsa, era agganciato alla stessa cella telefonica della vittima e spunta la figura di Damiano Guerinoni, che era uscito dalla discoteca da poco. Quando si cerca il match perfetto per i loci microsatellitari autosomici, non lo si ottiene né con il DNA di Damiano, né con quello dei fratelli o di altri familiari.

Si procede dunque alla ricerca della paternità dell'ignoto 1 e, con probabilità quasi del 100%, si arriva all'individuazione della figura paterna in Giuseppe Guerinoni, al momento deceduto, di cui però si era recuperata la saliva dai francobolli e di cui venne decisa la riesumazione della salma. Per quanto riguarda la maternità di ignoto 1, lo studio degli alleli obbligati materni con probabilità del 99.9% portò gli investigatori a Ester Arzuffi, dunque madre naturale dell'omicida. Emerse che la Arzuffi aveva avuto tre figli con il marito e uno (si seppe in seguito) fuori dal matrimonio, Daniele Bossetti, il

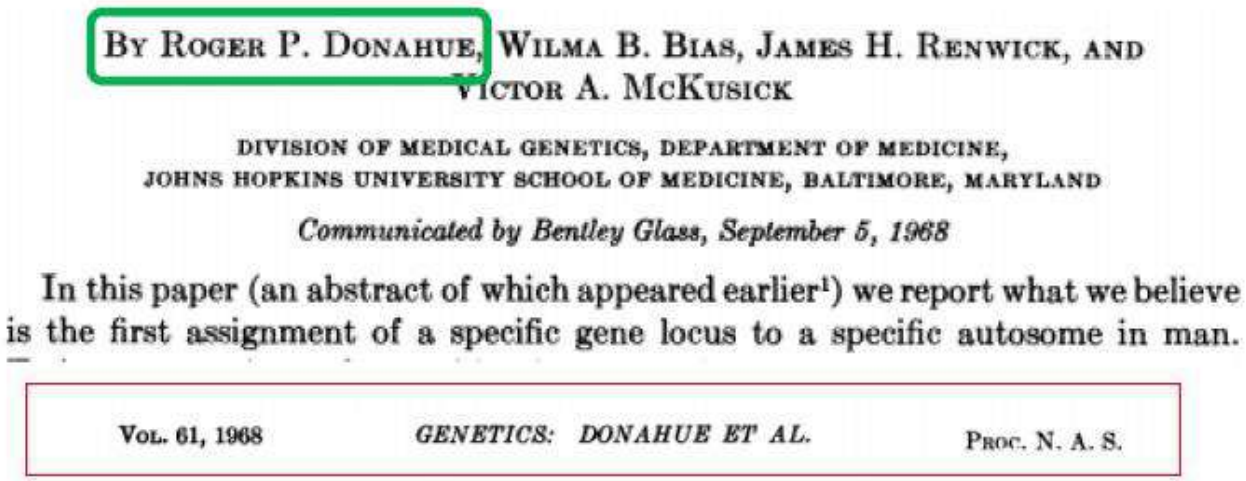
cui nome giunse agli inquirenti il 16 giugno 2014, in seguito ad un ordinario controllo alla circolazione attraverso alcol test, che consiste in un prelievo di saliva. Si osservò la completa compatibilità per 21 marcatori STR autosomici e solo 1 soggetto maschile su una popolazione di un miliardo di miliardi condivide tali caratteristiche genetiche. Questa è la dicitura giuridica adottata in questi casi: "Gli esiti di questa indagine genetica, dato l'elevatissimo numero delle ricorrenze statistiche confermative è tale da rendere infinitesimale la probabilità di un errore, presenta perciò natura di prova e non di mero riferimento indiziario ai sensi dell'art. 192, comma secondo, cod. proc. Pen., ecc.". Il cromosoma Y ha confermato un match perfetto con la linea di discendenza della famiglia Guerinoni, evidenziando un possibile caso di adulterio. Si organizzarono alcune ricerche in merito al lavoro dell'indagato e si venne a conoscenza che la sera dell'omicidio il cellulare di Yara e di Bossetti si agganciarono alla stessa cella. La madre ha sempre negato la relazione con Guerinoni ma ulteriori indagini genetiche sul presunto padre, dichiarato dalla madre, hanno dato esito negativo. La prova del DNA acquista sempre di più carattere probatorio. L'ergastolo viene ribadito al termine di tutti e tre i gradi di giudizio: il 16/06/2014 (processo di primo grado), il 17/07/2017 (in appello) e, infine, la Cassazione ha confermato la sentenza il 12/10/2018.

Il mappaggio dei geni

Mi chiedo spesso "perché mappare i geni"? Le ragioni sono molte, ma tre prevalgono su tutte. La prima ha a che fare con la sfida. Molti di noi erano spinti a mettere a punto metodiche di mappatura dei geni umani solo perché il problema sembrava a prima vista insuperabile. Una seconda motivazione è la soddisfazione estetica. Per qualche ragione ci gratifica ordinare un gene ben conosciuto nella sua posizione corretta sulla mappa. Tuttavia, c'è una terza ragione: quella utilitaristica. Credo di poter dire con certezza "il mappaggio dei geni è un bene per voi"!

FRANK RUDDLE (DA REVERSE GENETICS AND BEYOND. *AM. J. HUM. GENET. 1984; 36:944)*

Prima del 1950 era stato compiuto il mappaggio del cromosoma X, ottenuto attraverso la ricostruzione di pedigree, mentre il primo locus autosomico mappato fu descritto nel 1968, agli albori della citogenetica in un lavoro di McKusick, Donahue et al. pubblicato su "Genetics". Pubblicarono uno studio sul linkage del gruppo sanguigno Duffy al cromosoma 1.



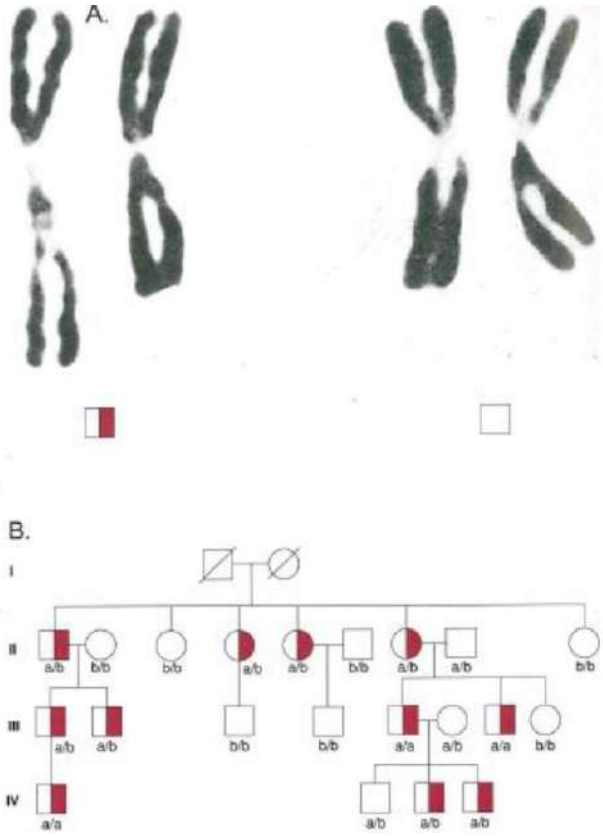
BY ROGER P. DONAHUE, WILMA B. BIAS, JAMES H. RENWICK, AND VICTOR A. MCKUSICK

DIVISION OF MEDICAL GENETICS, DEPARTMENT OF MEDICINE, JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE, BALTIMORE, MARYLAND

Communicated by Bentley Glass, September 5, 1968

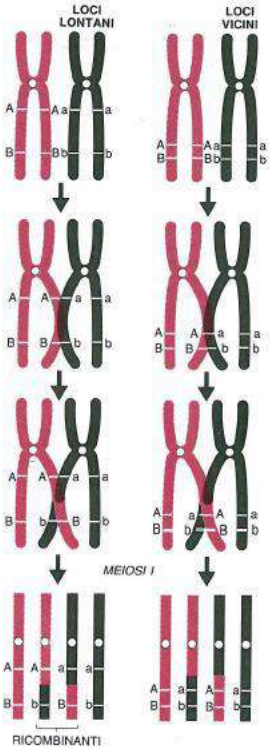
In this paper (an abstract of which appeared earlier[1]) we report what we believe is the first assignment of a specific gene locus to a specific autosome in man.

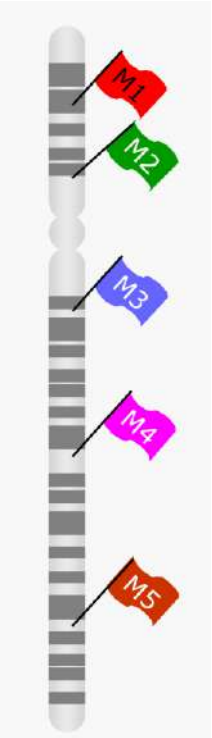
VOL. 61, 1968 GENETICS: DONAHUE ET AL. PROC. N. A. S.



Alcuni studi effettuati prima dell'introduzione delle tecniche di bandeggio cromosomico, rivelano un notevole incremento della lunghezza del cromosoma al di sotto del centromero, ritenuto un polimorfismo. I pedigree ricostruiti all'epoca mostrano la modalità di trasmissione del gruppo sanguigno Duffy e dell'eteromorfismo del cromosoma 1.

I simboli colorati indicano la presenza dell'eteromorfismo sul cromosoma 1, mentre a e b indicano il genotipo al gruppo sanguigno Duffy. Sebbene lo stato degli individui della generazione I non fosse noto, nel resto dell'albero si può notare che il cromosoma con l'eteromorfismo segrega cosegrega con l'allele a al locus Duffy. Il gene responsabile del gruppo sanguigno Duffy (FY) doveva trovarsi sul cromosoma 1.

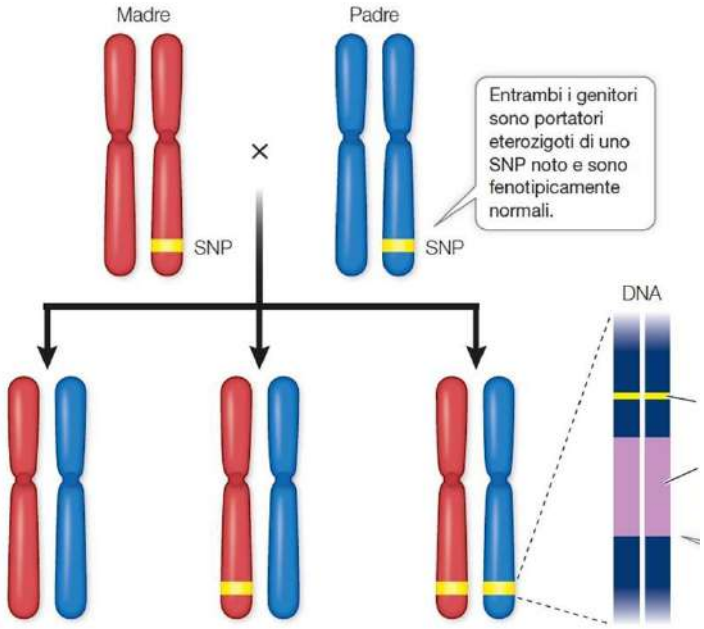




Questo fu il primo esempio di analisi genetica mediante linkage (concatenazione). Questo tipo di approccio consente di mappare un gene sulla base della sua vicinanza ad un altro locus sullo stesso cromosoma. Con le mappe genetiche degli organismi modello a disposizione degli scienziati all'epoca era possibile applicare questo metodo, ad esempio, su Drosophila, che mostrava fenotipi evidenti e che era oggetto di studio sui mutanti.
Sull'uomo ciò non valeva: emerse la necessità di trovare dei marcatori genetici facilmente identificabili. Si iniziarono a cercare loci genici che identificavano in maniera univoca una regione cromosomica. Fino a quel momento si erano usati marcatori morfologici, come caratteri mendeliani (eteromorfismo, forma della foglia, forma del pisello, ecc.), gruppi sanguigni, isoenzimi, proteine (del plasma, del latte, ecc.), adesso venivano sfruttati marcatori molecolari come RFLP, microsatelliti e SNP.

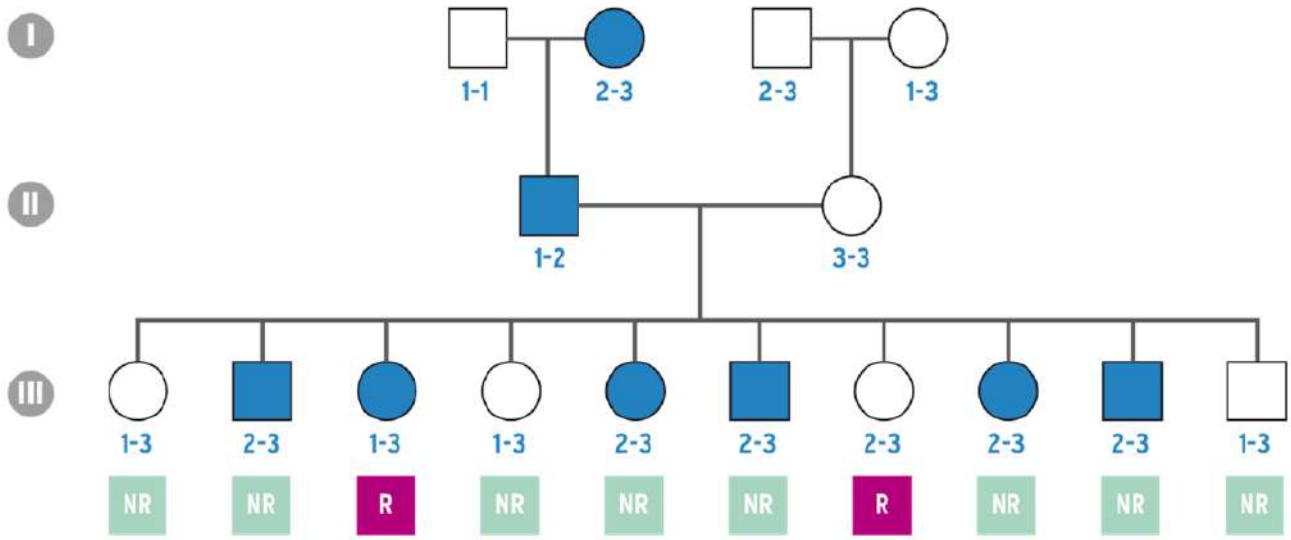
Nel 1994 su "Science" venne pubblicato un articolo dove si descriveva la prima mappa globale basata sul linkage nell'essere umano con densità espressa in centimorgan. Consisteva di 5840 loci, di cui

970 univocamente ordinati. Per i 2/3 si trattava di STR. La mappa aveva una densità media di 0,7 cM. Fu uno dei primi obiettivi raggiunti dell'HGP (Human Genome Project).



Un esempio di applicazione dei marcatori genetici è costituito dal gene tracking, il primo metodo diagnostico basato sul DNA, che utilizza le conoscenze sulla localizzazione di mappa del locus-malattia, ma non quelle sull'effettivo gene-malattia. I prerequisiti per il gene tracking sono: marcatori polimorfici strettamente associati al locus-malattia e campioni di DNA della famiglia. Il gene tracking implica tre passaggi logici:

1. Distinguere i due cromosomi del genitore;
2. Determinare la fase;
3. Determinare quale cromosoma sia stato trasmesso al probando.



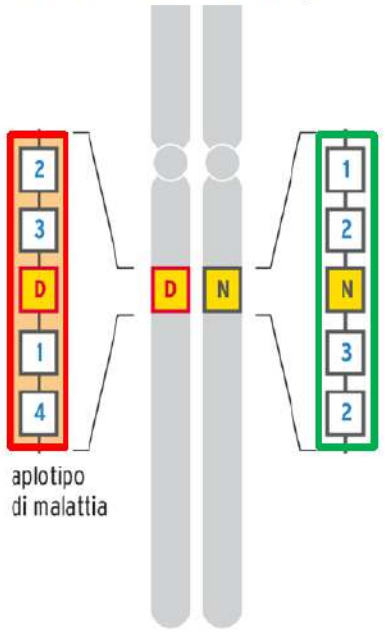
Si segue la cosiddetta strategia posizionale: si conosce solo la localizzazione cromosomica del locus. Richiede la costruzione di una mappa fisica e genetica della regione. Lo scopo è definire una regione candidata il più limitata possibile. E' importante disporre di numerosi siti polimorfici e con un alto grado di eterozigosità per costruire l'aplotipo di una famiglia. I primi dieci anni (dal 1986 al 1995) di strategia di tipo posizionale portarono all'identificazione dei geni responsabili della DMD, della fibrosi cistica, del retinoblastoma, della neurofibromatosi di tipo 1, della distrofia miotonica, della corea di Huntington, della malattia di von Hippel-Lindau, dell'acondroplasia, della malattia del rene policistico, dell'atrofia muscolare spinale e di molte altre ancora.

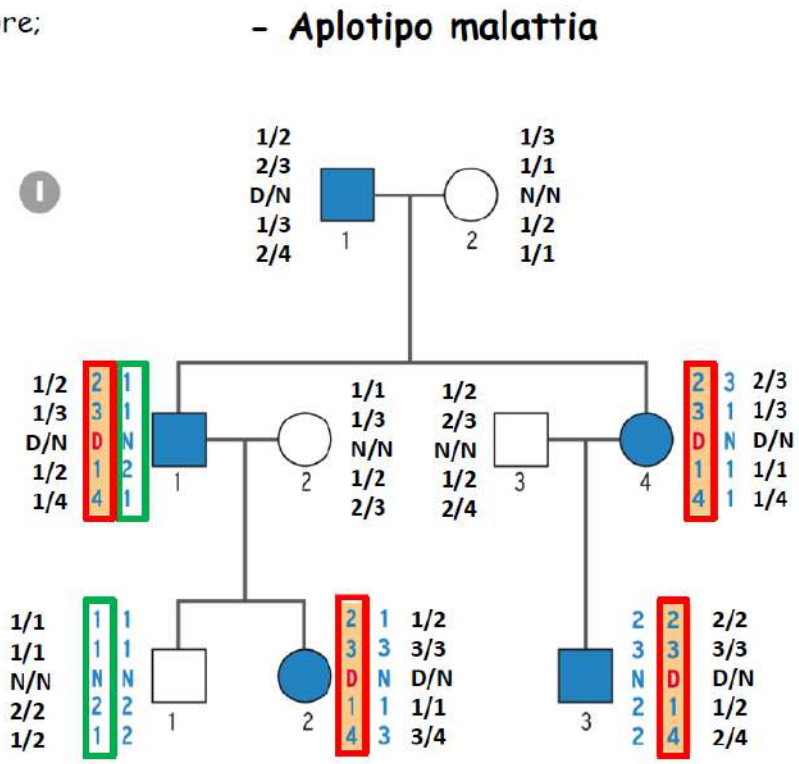
I primi dieci anni di strategie di tipo posizionale

Year	Disease	MIM number	Location	Gene	Chromosome abnormality
1986	Duchenne muscular dystrophy	310200	Xp21.3	DMD	(a) del(X)(p21.3) (b) t(X;21)(p21.3:p13)
	Retinoblastoma	180200	13q14	RB	del(13)(q13.1q14.5)
1989	Cystic fibrosis	219700	7q31	CFTR	None
1990	Neurofibromatosis 1	162200	17q11.2	NF1	Balanced translocations t(1;17)(p34.3:q11.2) t(17;22)(q11.2:q11.2)
	Wilms' tumor	194070	11p13	WT1	del(11)(p14p13)
1991	Aniridia	106210	11p13	PAX6	t(4;11)(q22;p13) del(11)(p13)
	Familial polyposis coli	175100	5q21	APC	del(5)(q15q22)
	Fragile-X syndrome	309550	Xq27.3	FMR1	FRAXA fragile site
	Myotonic dystrophy	160900	19q13.3	DMPK	None
1993	Huntington's disease	143100	4p16	HD	None
	Tuberous sclerosis 2	191092	16p13	TSC2	Microdeletions in candidate region
	von Hippel-Lindau disease	193300	3p25	VHL	Microdeletions in candidate region
1994	Achondroplasia	100800	4p16	FGFR3	None
	Early-onset breast/ovarian cancer	113705	17q21	BRCA1	None
	Polycystic kidney disease	173900 601313	16p13.3	PKD1	t(16;22) (p13.3;q11.21)
1995	Spinal muscular atrophy	253300 600354	5q13	SMN1	None

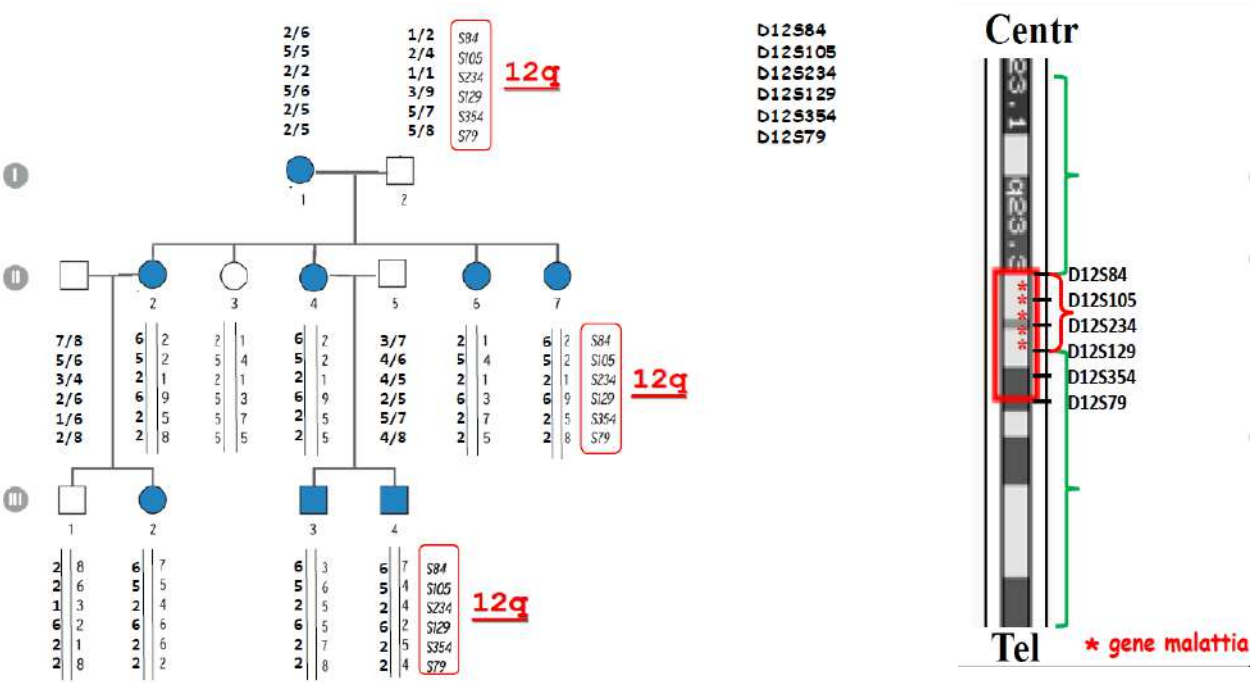
The data illustrate how important cytogenetic abnormalities were for cloning disease genes. For TSC2 and VHL, microdeletions were identified only late in the refinement of the candidate region.

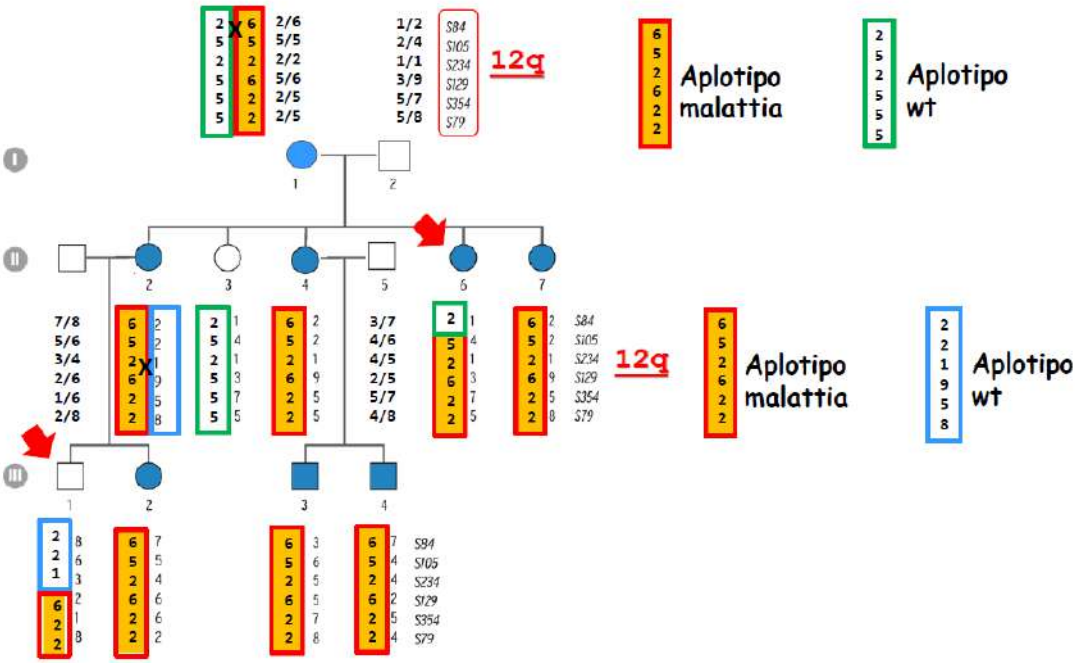
a) Distinguere i due crom del genitore;
b) Determinare la fase;
c) Determinare quale cromosoma sia stato trasmesso alla prole.

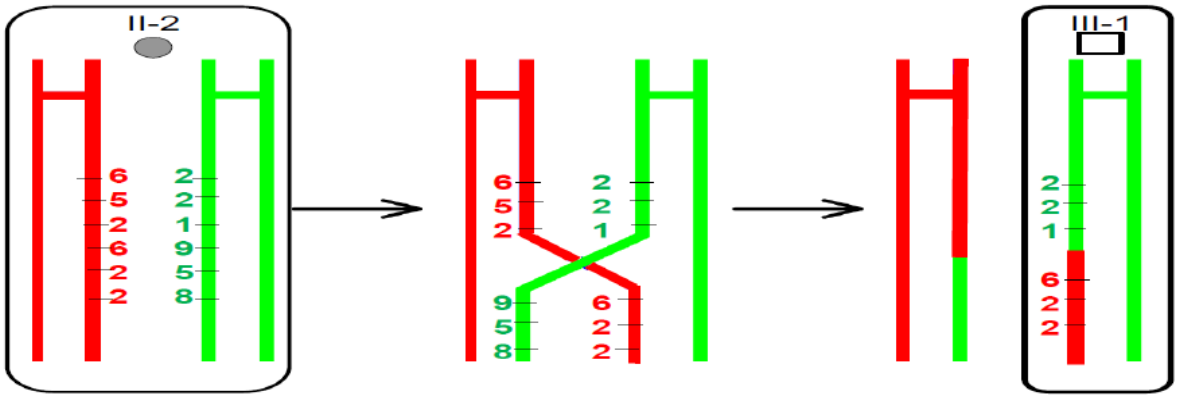




Lo studio realizzato da T.Strachan et al. nel 1994 per la ricerca del gene responsabile della malattia di Darier-White (un disordine della pelle) si basava su ricerche condotte anche da altri laboratori che avevano individuato la regione responsabile della malattia con marker che si legavano tra 12q23.1 e q24.1. Il lavoro di Strachan et al. definì fenomeni di crossing-over che permisero di restringere la posizione del gene malattia tra i marker D12S105 e D12S129.


12q
D12S84
D12S105
D12S234
D12S129
D12S354
D12S79
Centr
Tel
* gene malattia


12q
Aplotipo malattia
Aplotipo wt
12q
Aplotipo malattia
Aplotipo wt


II-2
III-1

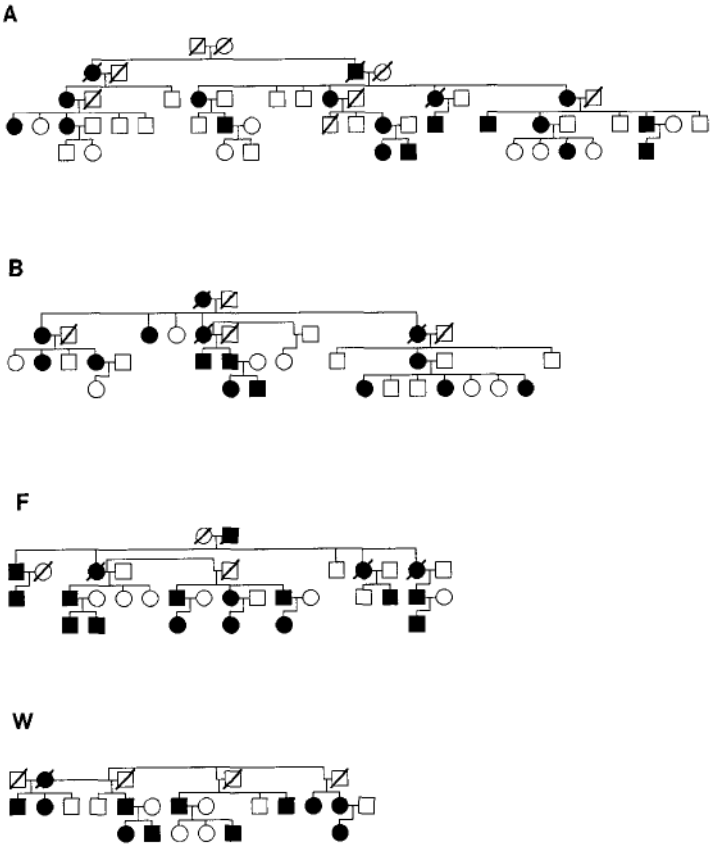
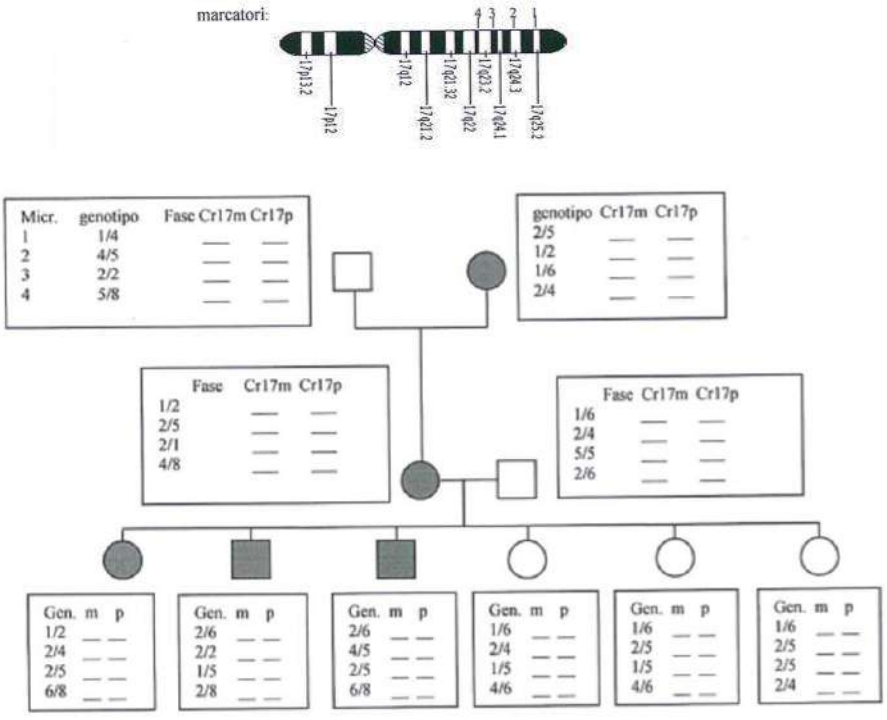
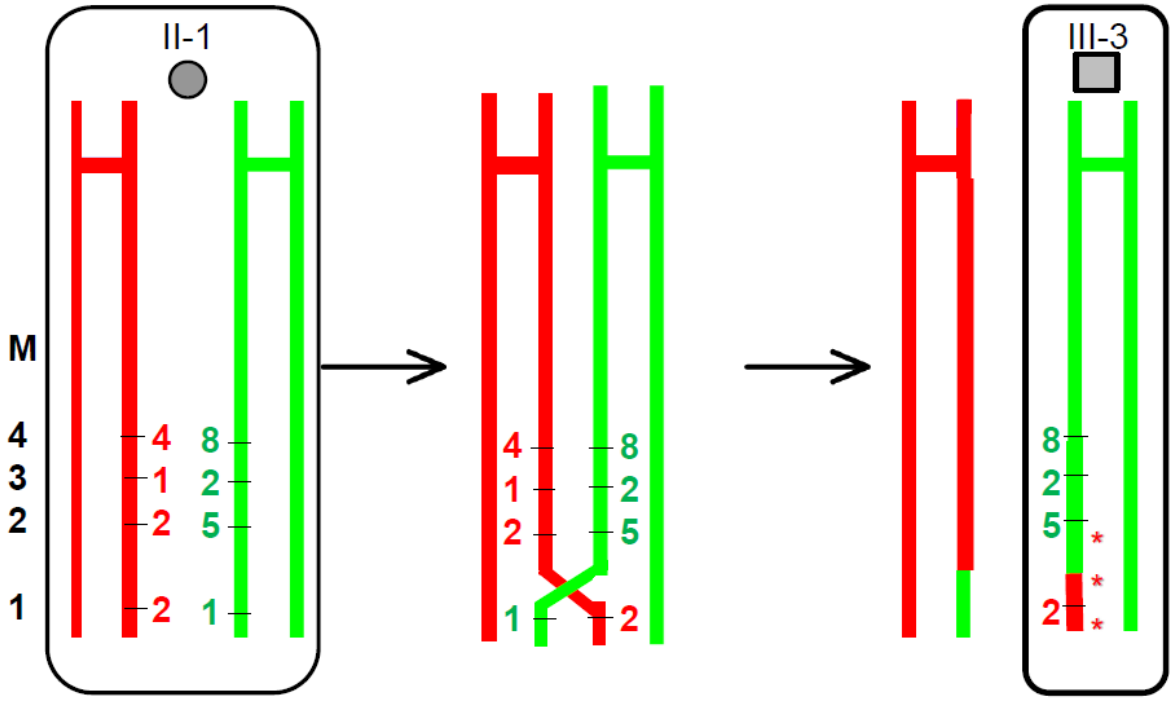


FIG. 1. Darier disease pedigrees. The pedigrees shown are condensed versions of known kindreds A, B, F, and W, showing in addition to deceased family members only those family members from which samples were analyzed. Pedigrees A and B from the northeast of England have been described previously (1). Pedigrees F and W are from North Yorkshire and Glasgow, respectively; in each of these families three members with Darier disease also had epilepsy, and in each family a fourth case also had obvious mental retardation. Affected status was determined by the presence of one or more of the following four clinical features: the typical rash, pitting of the palmar sweat ridges, longitudinal striations of the nails with distal notching, and multiple large or small plane wart-like lesions of the dorsal hands. In practice the distinction was clear; all cases identified had two or more of those features, with most having all four (9). Penetrance is complete when adults are ascertained in this way (9), but as this is not the case in childhood, unaffected subjects under the age of 16 were not studied.

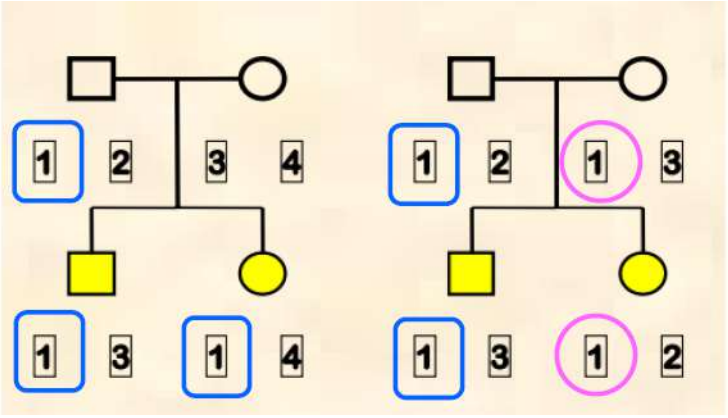
Esercizio: Indicare la regione candidata del cromosoma 17 che potrebbe contenere il gene responsabile di un carattere autosomico dominante.



* gene malattia



Per alcune famiglie, di consanguinei, è possibile compiere una mappatura per autozigosi. Essa indica l'omozigosi di alleli identici per discendenza in individui nati dall'unione fra consanguinei. Se due individui sono parenti, essi hanno almeno un antenato comune, quindi vi è una certa probabilità che un allele portato da uno di loro e un allele portato dall'altro siano entrambi discendenti dalla stessa molecola di DNA. Gli individui con malattie recessive in famiglie di consanguinei possono presentare autozigosi per i geni associati alla malattia. Vanno anche introdotti i concetti di IBS (Identical By State) e IBD (Identical By Descent). Prendiamo in esame i due seguenti alberi genealogici.



Nell'albero di sinistra, l'allele 1 si presenta nel padre e nei due figli, dunque siamo in una situazione di IBS, perché, se dovessimo scrivere i genotipi di quei tre individui, avremmo: 1-2, 1-3, 1-4- In tutti e tre compare l'allele con 1 repeat. E' anche un contesto di IBD, perché l'allele 1 ai figli può averlo trasmesso soltanto il padre, dal momento che la madre ha genotipo 3-4.

Osserviamo adesso il pedigree di destra: anche in questo caso entrambi i figli hanno ereditato l'allele 1 da uno dei genitori, quindi è comunque un IBS. Non si tratta di IBD, però, perché il figlio maschio ha ereditato il proprio 1 dal padre, mentre la figlia dalla madre, mentre prima era stato il padre a fornire

l'allele 1 a entrambi. Ce ne si accorge facilmente prendendo in esame il genotipo del figlio maschio: l'allele 3 lo deve aver ereditato per forza dalla madre, dunque l'1 proviene dal padre. Mentre la femmina ha l'allele 2 su un cromosoma (discendenza paterna), perciò l'1 è di eredità materna.

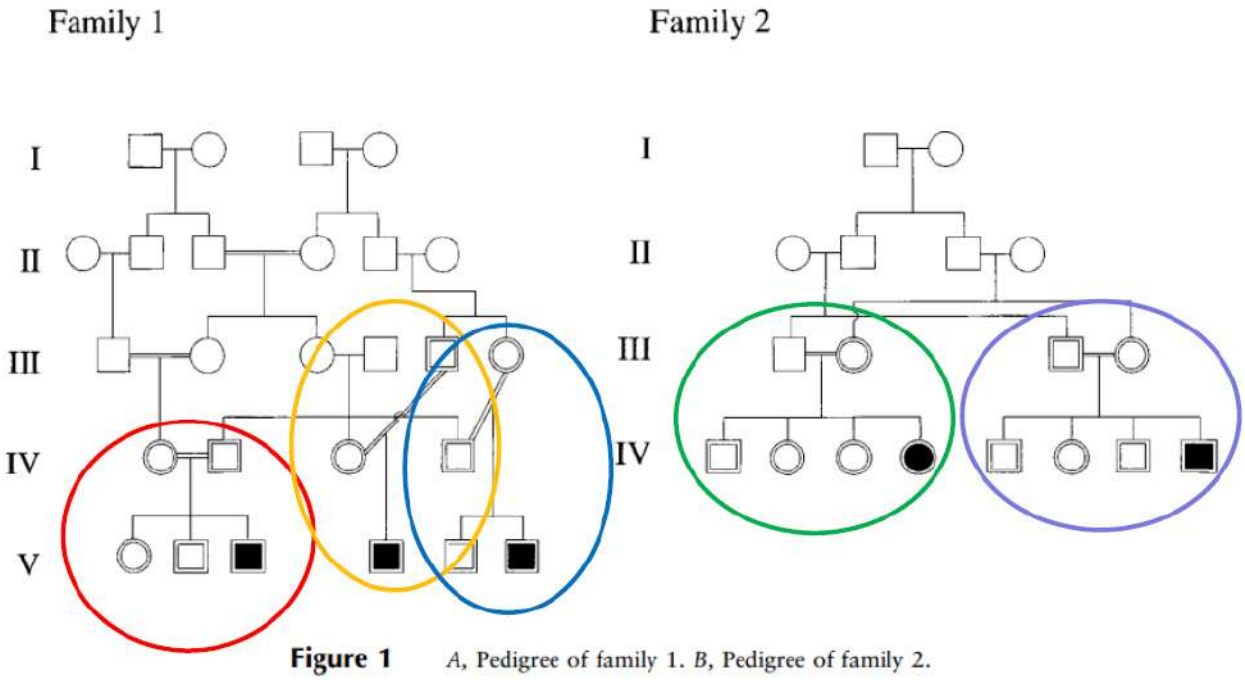


Figure 1 *A*, Pedigree of family 1. *B*, Pedigree of family 2.

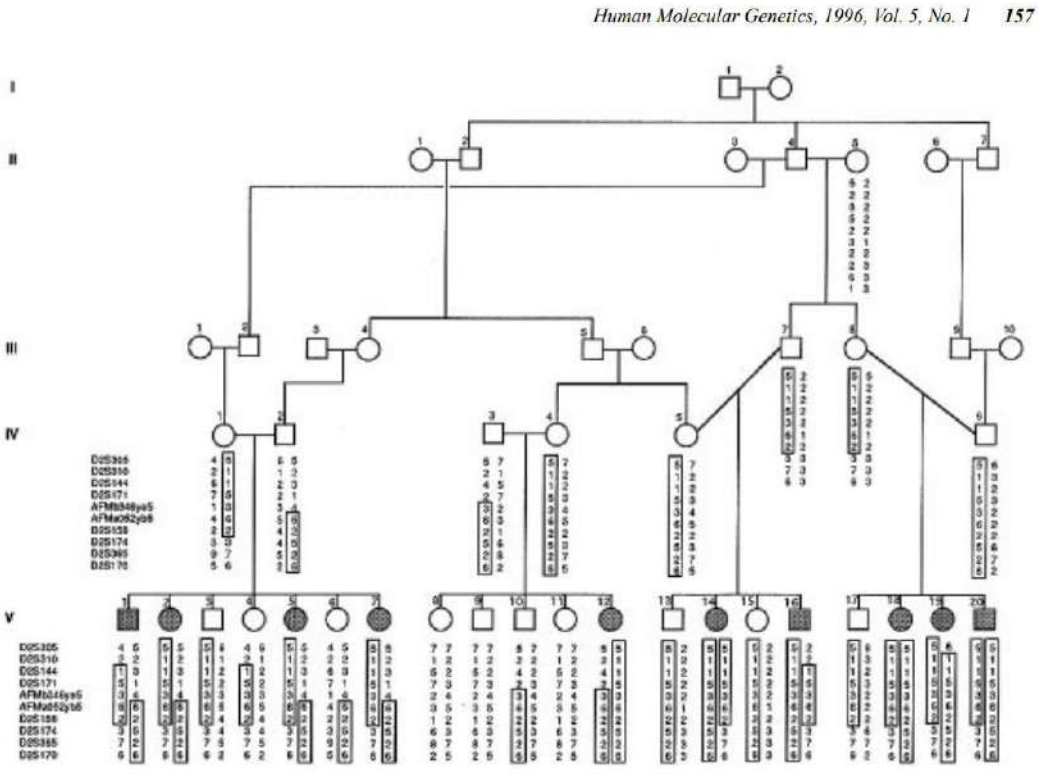


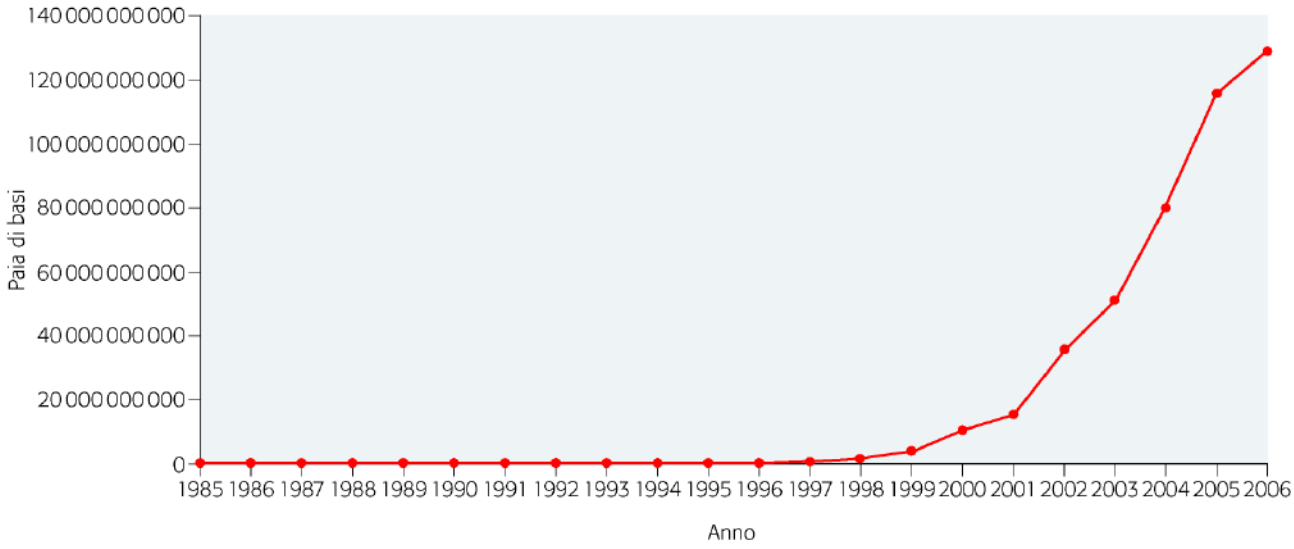
Figure 1. Family F, showing segregation of the linked loci. The haplotype associated with *DFNB6* in each individual is boxed. Dark symbols represent deaf individuals.

Uno studio basato sull'autozigosi venne pubblicato nel 1995 su "Human Molecular Genomics" e riguardava un gene responsabile della sordità sensorineurale non sindromica recessiva, che si scoprì mappare sul

cromosoma 2, braccio lungo, regione 22-23. Venne descritto un intervallo massimo di 2 cM per il gene DFNB6, attraverso l'osservazione di ricombinanti e l'ausilio di mappaggi per omozigosi.

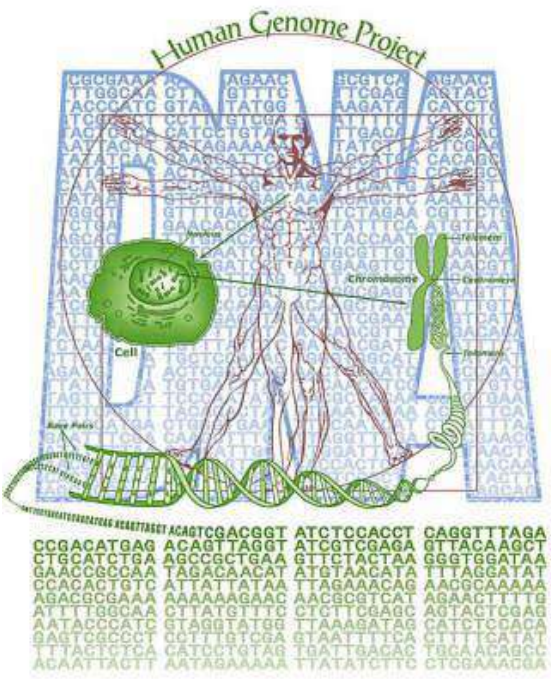
Il sequenziamento del genoma umano

Organism	Completion date	Size	Description
phage phiX174	1978	5,368 bp	1st viral genome
human mtDNA	1980	16,571 bp	1st organelle genome
lambda phage	1982	48,502 bp	important virus model
HIV	1985	9,193 bp	AIDS retrovirus
H. influenzae	1995	1,830 Kb	1st bacterial genome
M. genitalium	1995	580 Kb	smallest bacterial genome
S. cerevisiae	1996	12.5 Mb	1st eukaryotic genome
E. coli K12	1997	4.6 Mb	bacterial model organism
C. trachomatis	1998	1,042 Kb	internal parasite of eukaryotes
D. melanogaster	2000	180 Mb	fruit fly, model insect
A. thaliana	2000	125 Mb	thale cress, model plant
H. sapiens	2001	3,000 Mb	human
SARS	2003	29,751 bp	coronavirus



Il primo organismo di cui venne sequenziato interamente il genoma fu il batterio Haemophilus influenzae, nel 1995. Seguirono, nel 1996, il genoma del lievito (Saccharomyces cerevisiae), nel 1998 un Nematode (Caenorhabditis elegans) e nel 2000 quello del moscerino della frutta (Drosophila melanogaster).

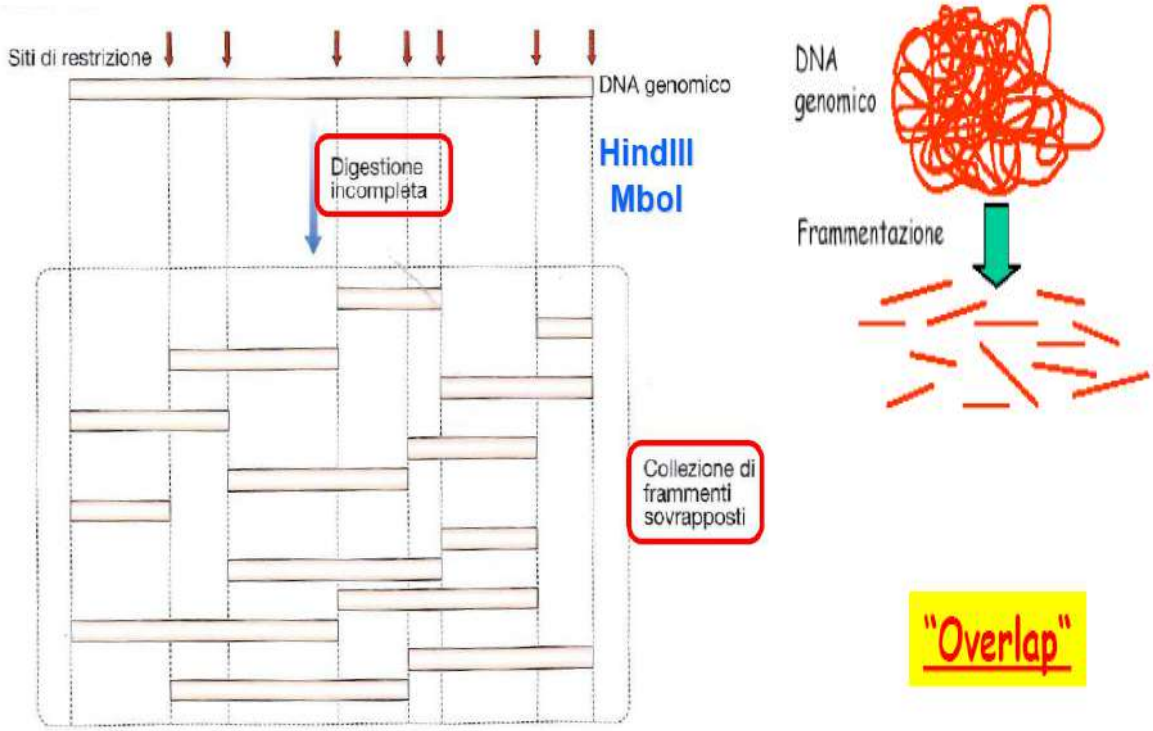
Nel 1984 venne lanciata l'idea, da parte di Robert Sinsheimer, di dar vita ad un progetto con l'ambizioso obiettivo di sequenziare l'intero genoma umano. Dal 1984 al 1986 il progetto venne promosso dall'allora direttore del Dipartimento dell'Energia Charles de Lisi, che chiese e ottenne i fondi necessari dal governo americano. Nel 1983, intanto, era stato fondato Genbank. Il 7 marzo 1986 Renato Dulbecco, illustre scienziato italiano, Premio Nobel per la Medicina nel 1975, pubblicò



un articolo su "Science" in cui sottolineava l'importanza del Progetto Genoma Umano per la ricerca nella lotta contro il cancro, per comprendere meglio i meccanismi che ne stavano alla base. Lo stesso anno De Lisi organizzò un workshop internazionale e negli anni successivi ('87-'88) finanziò la ricerca con provenienti dal budget del proprio dipartimento. Nel 1988 il US National Research Council e il US National Institute of Health (NIH) concentrarono l'attenzione sugli organismi modello come E.coli, Drosophila, C. elegans. Nel 1990 ebbe finalmente inizio lo Human Genome Project, in collaborazione tra DOE (Department of Energy) e NIH, sotto la guida di James Watson, Direttore dell'Ufficio per la Ricerca. In seguito si unirono altri paesi, come il Regno Unito, la Francia, il Giappone, la Germania, la Cina e la Russia.

Il piano d'azione dell'HGP prevedeva un tempo totale stimato di 15 anni (1990-2005), il coinvolgimento di 5 Paesi, riuniti nell'IHGSC (International Human Genome Sequencing Consortium) e una spesa che sarebbe dovuta ammontare intorno ai 3 miliardi di dollari, circa 1 dollaro a paio di basi.

Il DNA fu sottoposto a frammentazione o attraverso una rottura meccanica o con l'azione "digestiva" incompleta degli enzimi di restrizione come HindIII o MboI. I frammenti raccolti vennero poi raccolti in una collezione dove si osservò l'overlap, cioè la sovrapposizione di alcuni segmenti su altri.



Il passo successivo fu l'allestimento di una libreria di cloni con inserti lunghi che rappresentava il genoma 10x, dove 10x era il cosiddetto "fattore di copertura".

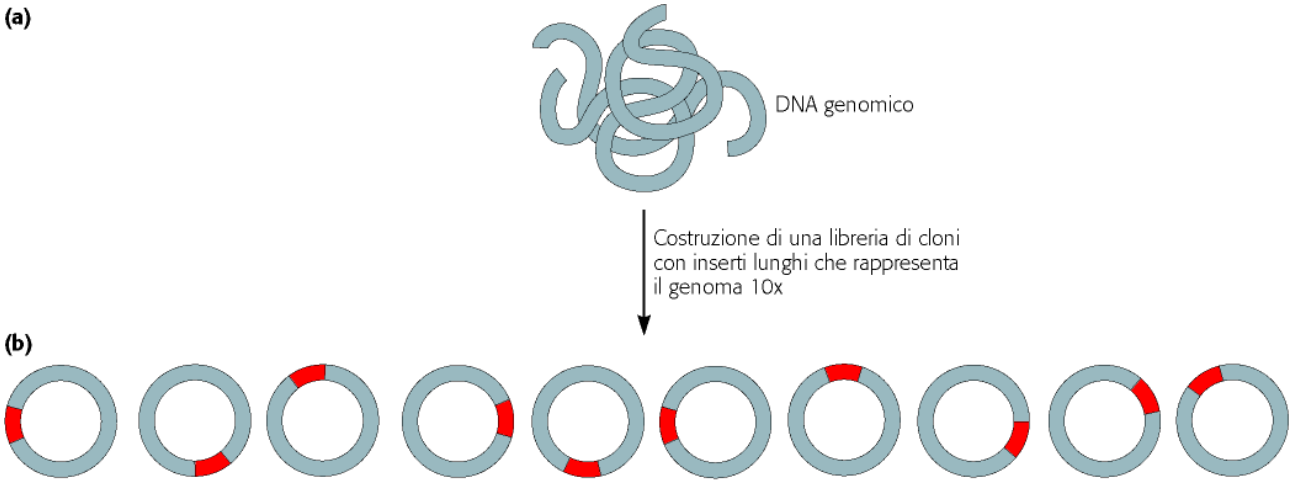
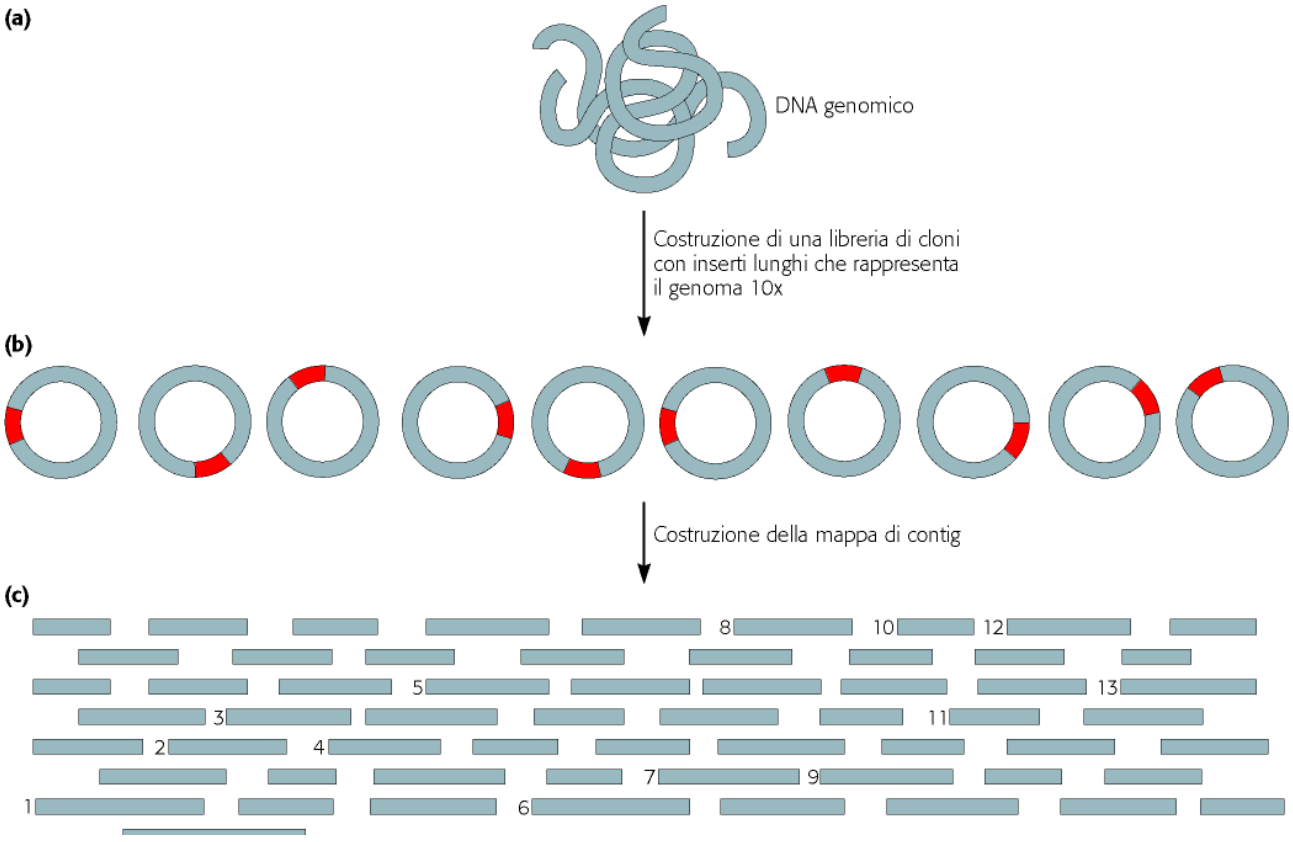


TABLE 7-2 Approximate Maximum Length of DNA That Can Be Clones in Vectors

Vector Type	Cloned DNA (kb)
Plasmid	20
λ phage	25
Cosmid	45
P1 phage	100
BAC (bacterial artificial chromosome)	300
YAC (yeast artificial chromosome)	1000

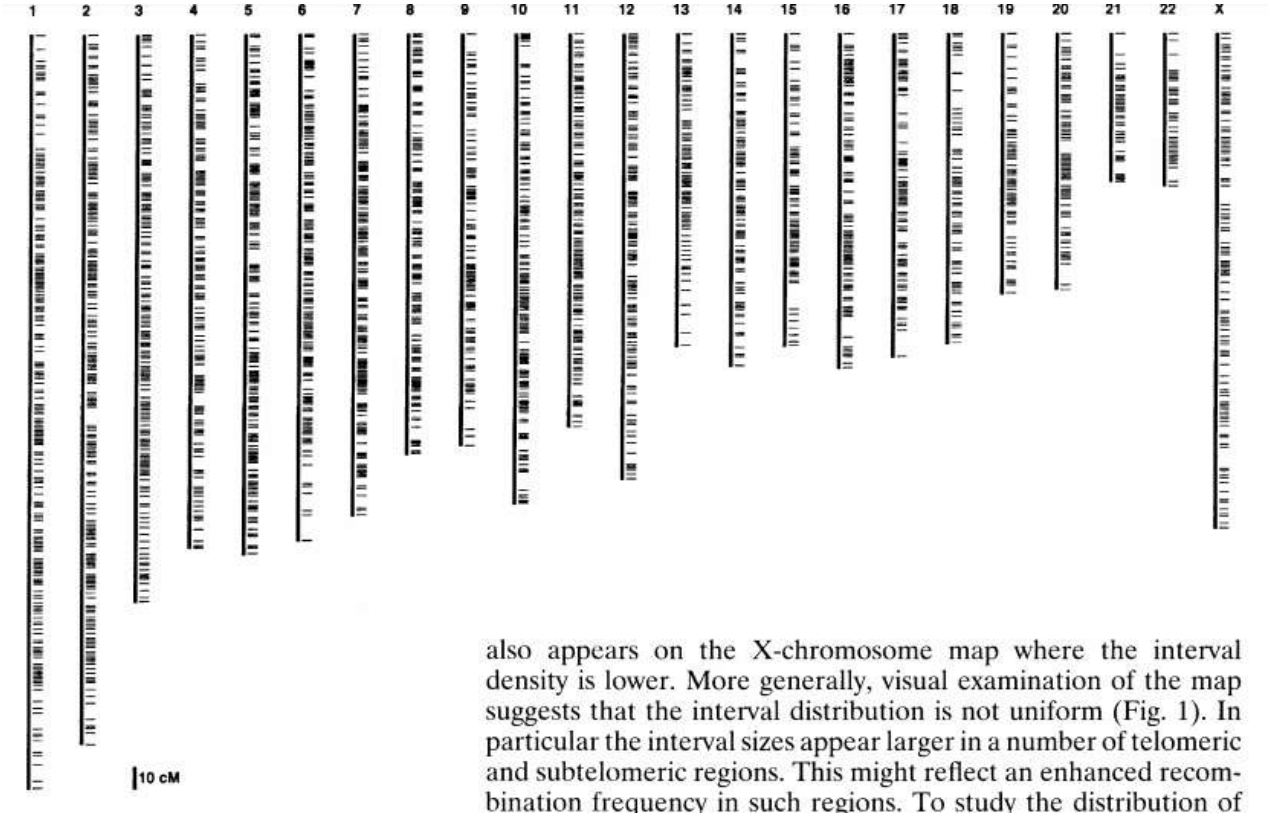
Come vettori di clonaggio vennero selezionati i BAC, dal momento che permettevano il clonaggio di frammenti piuttosto estesi, di circa 300 kb.

Venne eseguito il mappaggio dei cloni e si ricorse ai marcatori molecolari già visti in precedenza (RFLP, STR, SNPs).



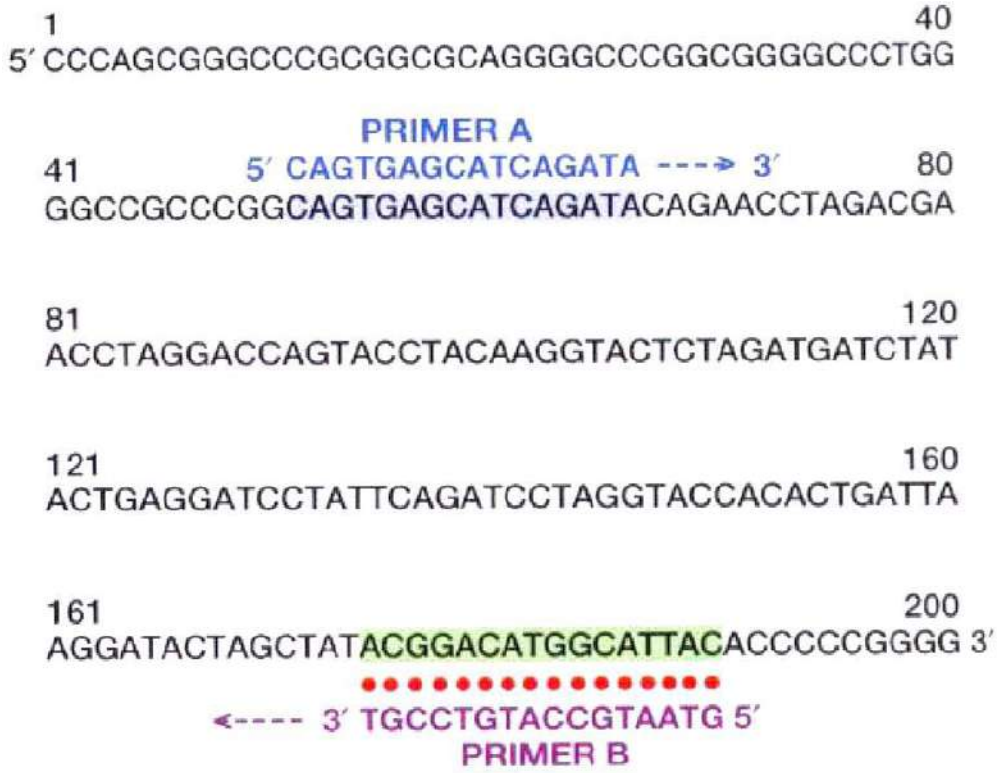
Nel 1996, su "Nature", venne pubblicato uno studio che, alla luce dei progressi compiuti nella mappatura su base linkage, riportava una mappa generale del genoma umano con l'ausilio di 5 264

microsatelliti. L'intervallo medio misurava 1,6 cM; il 59% della mappa era coperto da intervalli di 2 cM al massimo, mentre il restante 41% era dato da intervalli superiori ai 10 cM.

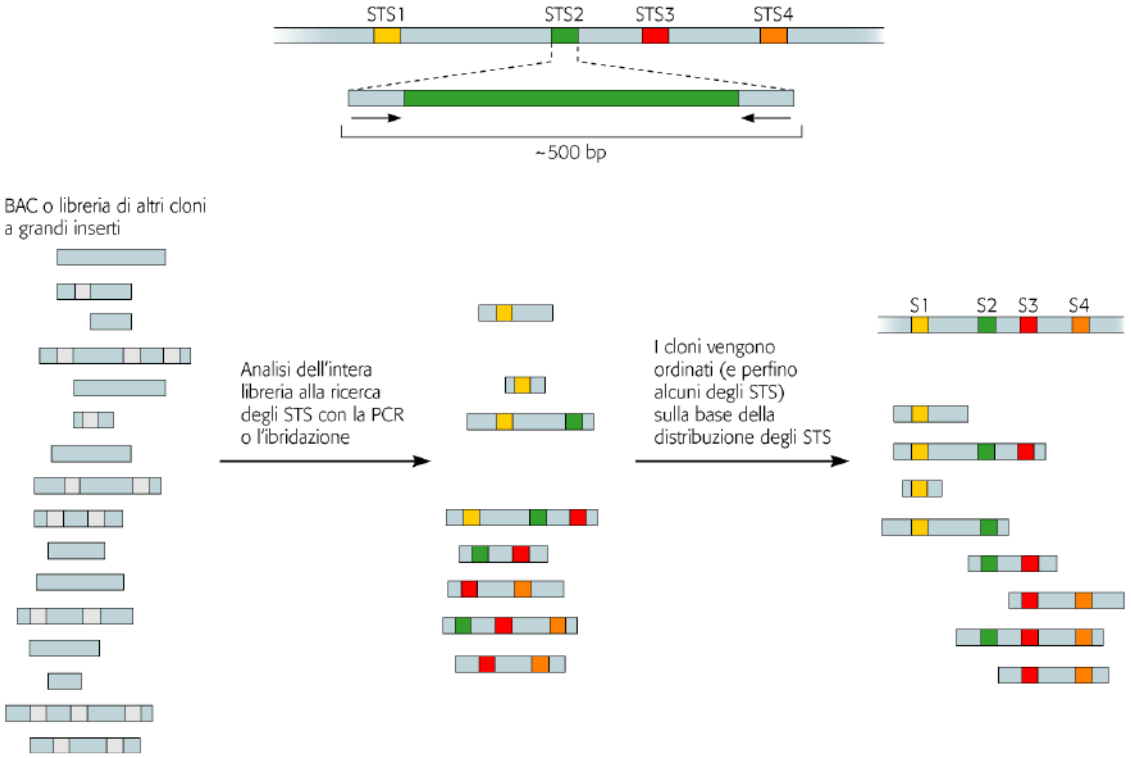


also appears on the X-chromosome map where the interval density is lower. More generally, visual examination of the map suggests that the interval distribution is not uniform (Fig. 1). In particular the interval sizes appear larger in a number of telomeric and subtelomeric regions. This might reflect an enhanced recombination frequency in such regions. To study the distribution of

Per il mappaggio dei cloni si scelse il metodo dei STS, i siti a sequenza etichettata, porzioni fisiche di DNA di 200-500 bp, con sequenza unica e nota.



Quindi si analizzò l'intera libreria di cloni BAC alla ricerca degli STS con la PCR e, infine, i cloni vennero ordinati (e perfino alcuni degli STS) sulla base della distribuzione degli STS.



Questo processo può determinare in modo non ambiguo l'ordine fisico dei cloni BAC, questo perché l'STS identifica inequivocabilmente la molecola di DNA più grande che lo contiene.

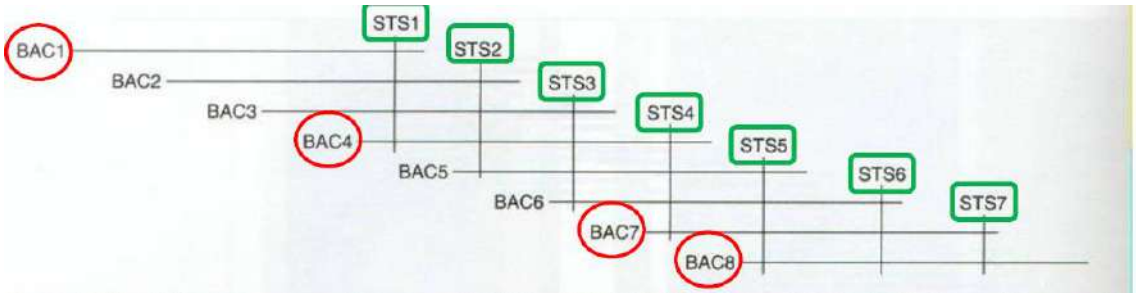
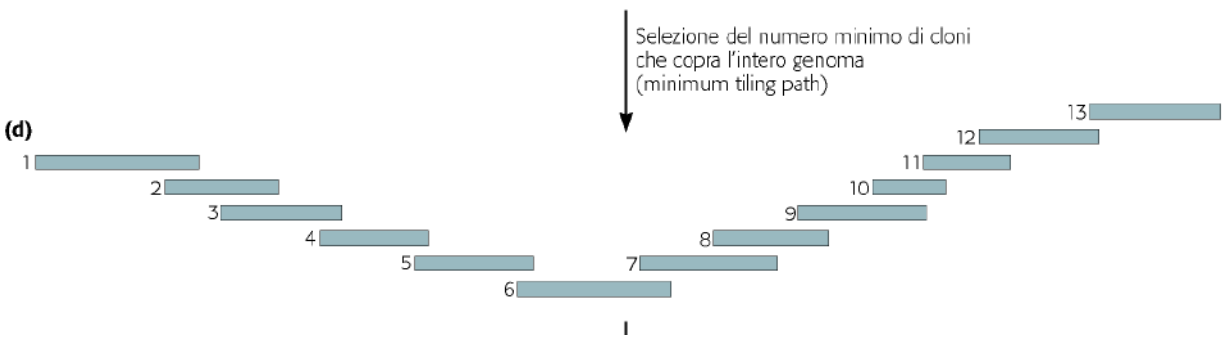


Figura 10.11 Un'ipotetica mappa fisica generata dall'analisi di siti di sequenze-etichetta (STS). Ciascun STS rappresenta u segmento unico del genoma che è stato amplificato per PCR da una coppia di primer. Qui, otto cloni BAC sono analizzati con sette marca In questo esempio, questo processo può determinare in modo non ambiguo l'ordine fisico dei cloni BAC.

Lo step successivo prevedeva il minimum tiling path (MTP), che è la selezione di un numero minimo di cloni in grado di coprire un intero genoma.



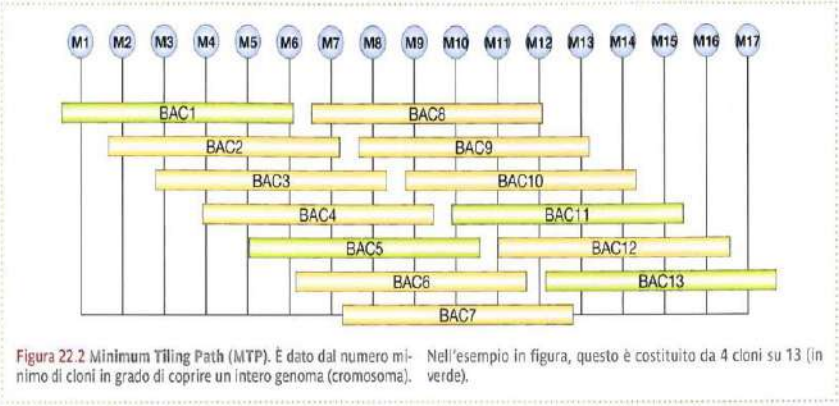
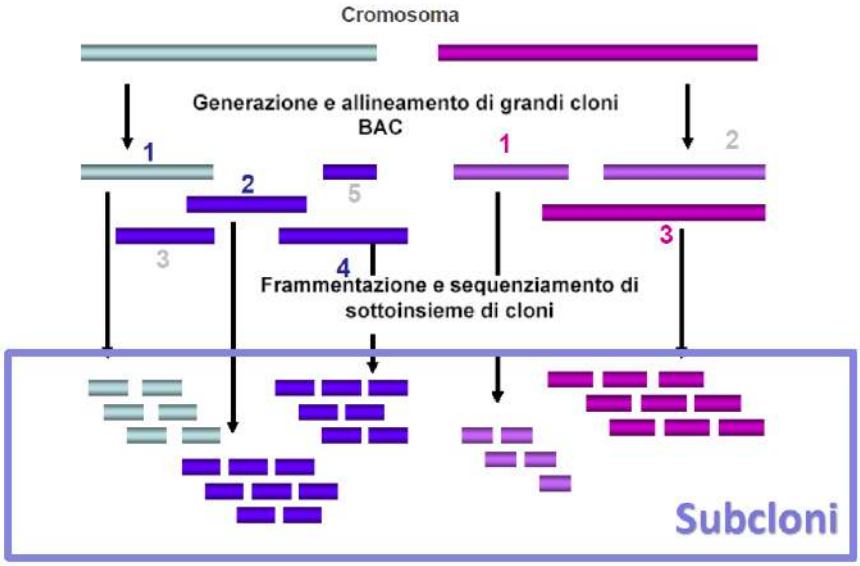
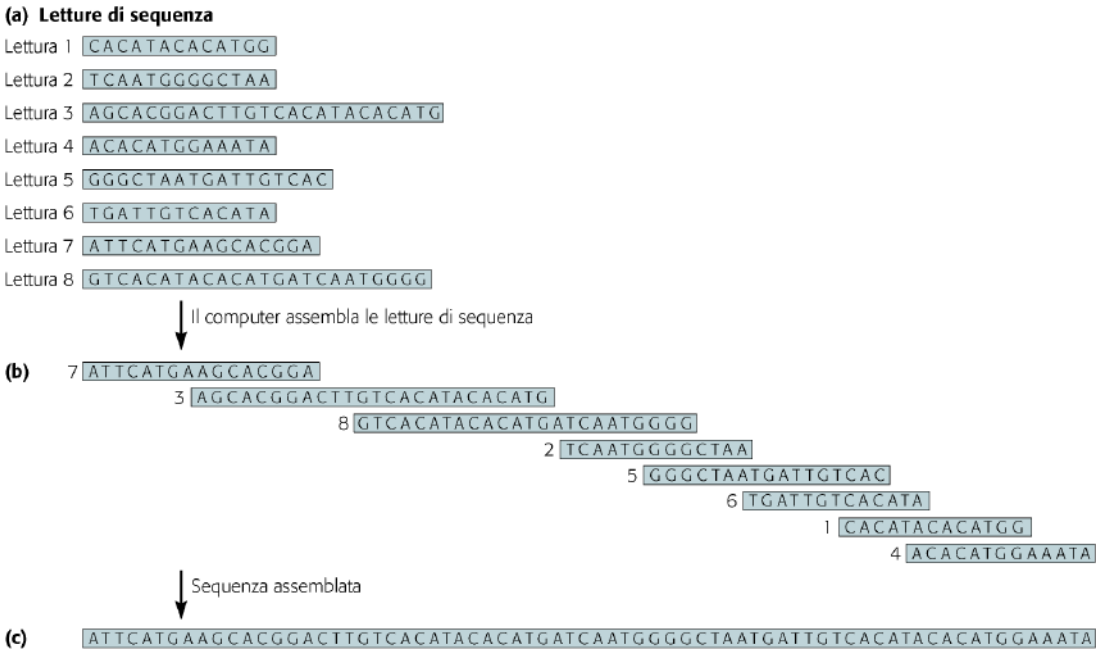


Figura 22.2 Minimum Tiling Path (MTP). È dato dal numero minimo di cloni in grado di coprire un intero genoma (cromosoma). Nell'esempio in figura, questo è costituito da 4 cloni su 13 (in verde).

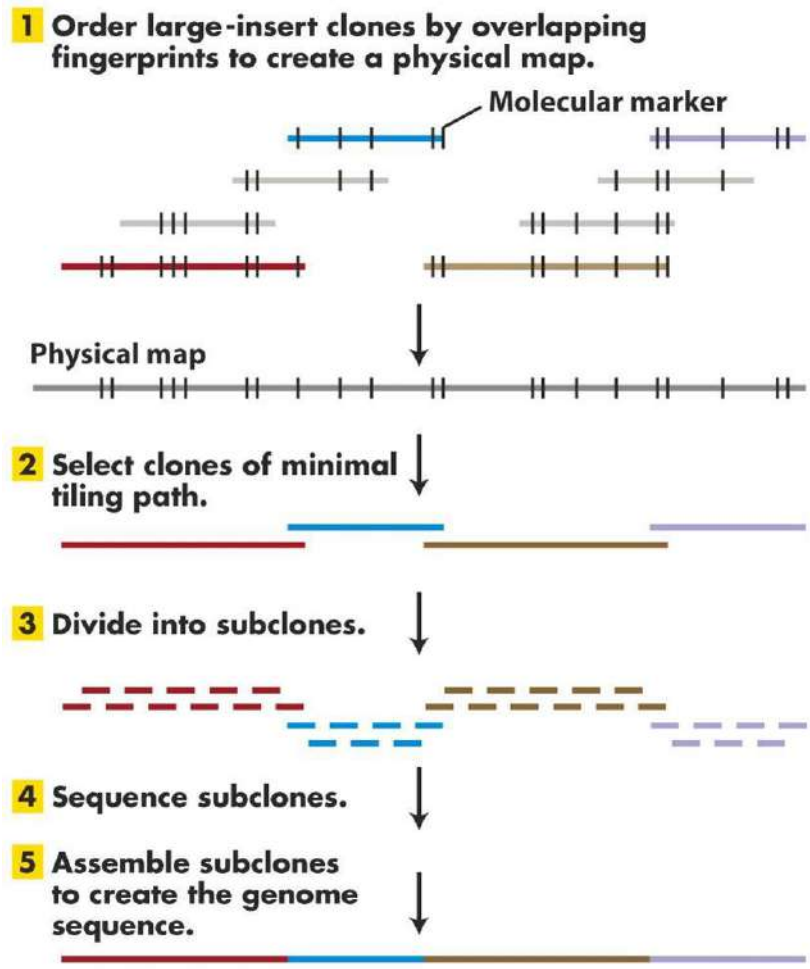
I cloni BAC allineati vennero dunque frammentati in tanti subcloni, che a loro volta vennero sequenziati con apparecchi che sfruttano l'elettroforesi capillare (a 1, 16 o 96 capillari).

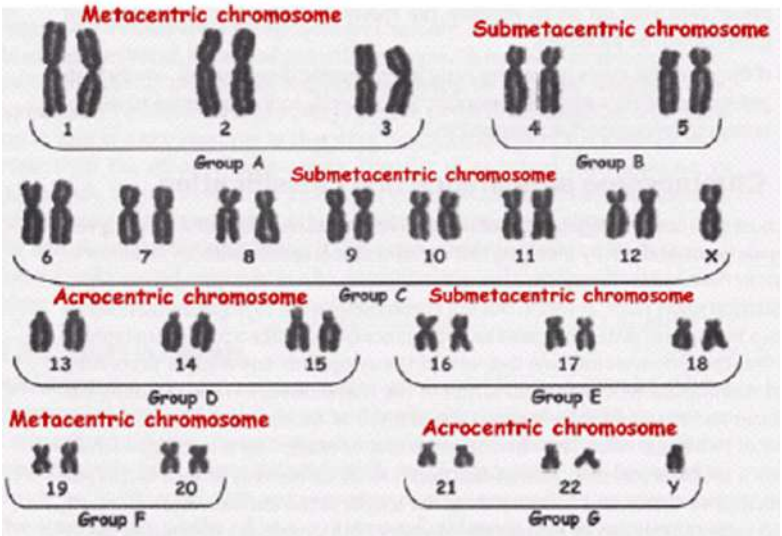


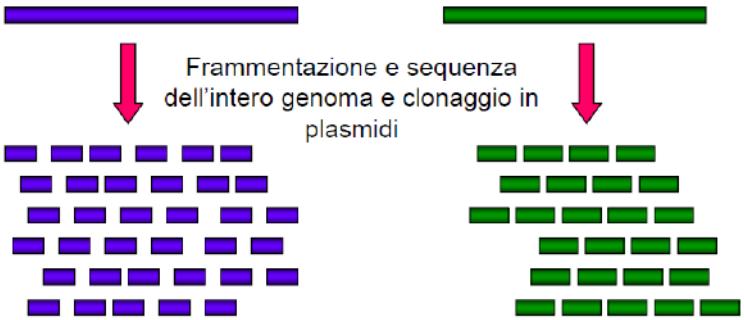
L'ultimo passaggio consisteva nell'assemblaggio delle "letture" di sequenze sovrapposte.



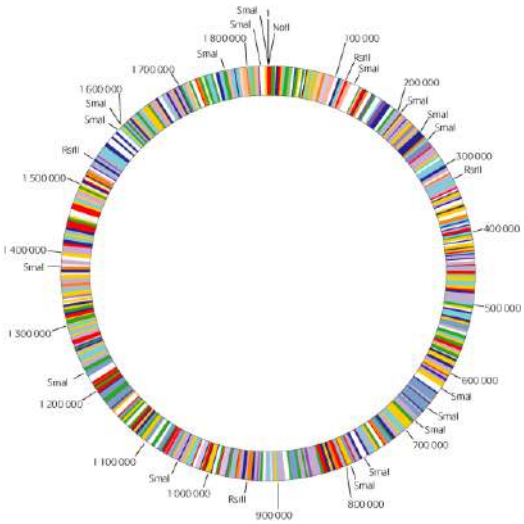
Così si operò per il sequenziamento di S. cerevisiae e C. elegans. I principi delle Bermuda enunciati nel 1996 stabilirono che i risultati del Progetto Genoma Umano dovessero essere resi pubblici e fruibili a tutti. Nei primi 8-9 anni venne sequenziato il cromosoma 22, il primo cromosoma ad essere sequenziato. Si attivarono centri nei 6 Paesi partecipanti, i più attivi furono Regno Unito e Stati Uniti. Si seguì un sequenziamento di tipo "gerarchico", detto anche "CGS", Clone by clone Genome Sequencing, o "top-down"), proposto da F.Collins.



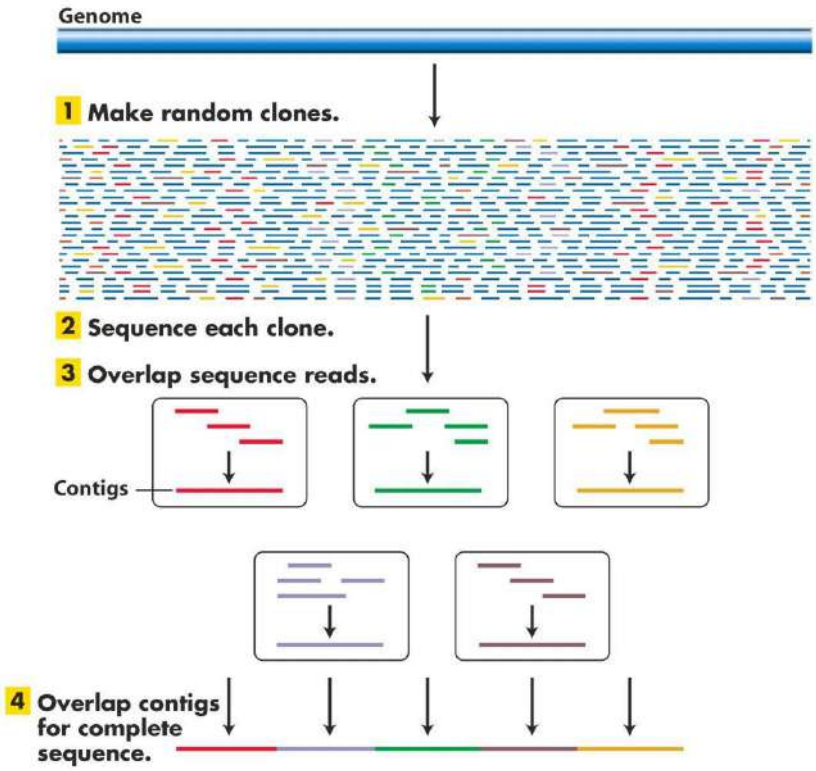




Un metodo alternativo venne ideato da C.Venter,: il sequenziamento "shotgun", o "WGS", Whole Genome Shotgun Sequencing, o "bottom-up". Venne pensato agli inizi del '90 da Craig Venter, che lavorava presso l'Institute for Genomic Research nel Maryland, per il sequenziamento dei batteri. In questo modo, ad esempio, era stato sequenziato il genoma di Haemophilus influenzae. Il grande vantaggio era il fatto che si potesse procedere senza possedere alcuna conoscenza a priori, senza avere alcuna mappa a disposizione.



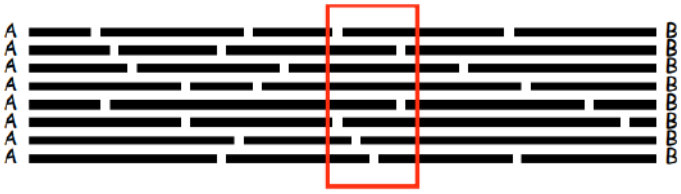
Questa strategia di sequenziamento funzionava su genomi piccoli e privi, o quasi, di ripetizioni, come quelli dei microrganismi, dei quali non si disponeva di mappe genetica. Nel 2006, ad esempio vennero sequenziati 300 organismi procariotici con dimensione variabile tra 1 e 3 Mb. Nel '98 Venter annunciò la nascita di una nuova "company", la Celera Genomics, che avrebbe sequenziato l'intero genoma in 2-3 anni.



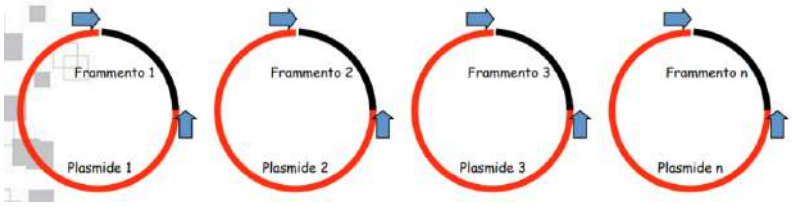
La "proof of principle", la dimostrazione della validità del metodo proposto, fu il sequenziamento di Drosophila melanogaster, che possiede un DNA di circa 180 Mb, pubblicato su "Science" nel 2000. Per questo fine venne reclutato un team di informatici per lo sviluppo di nuovi algoritmi, che permise il sequenziamento di circa 1,5 milioni di subcloni. Dimostrò che era possibile applicare la stessa metodica ad un genoma 20 volte più grande: in più di un anno, assemblò la sequenza genomica umana.

Questo era l'ordine delle operazioni: frammentazione in subcloni, clonaggio in plasmidi e sequenziamento di ciascun subclone, sovrapposizione delle letture di sequenze e sovrapposizione di sequenze contigue per ottenere la sequenza completa. Il metodo shotgun frammenta "semplicemente" il genoma in piccole unità sequenziabili e si affida a potenti algoritmi per ricostruire l'ordine dei frammenti. Nel sequenziamento shotgun si usano algoritmi per ricostruire la sequenza dei contig derivati da migliaia di cloni sovrapposti (Celera ne ha usati 5 differenti). Come nel sequenziamento gerarchico lo scopo è quello di ottenere una ridondanza di ciascun frammento del genoma da 5 a 10 volte.

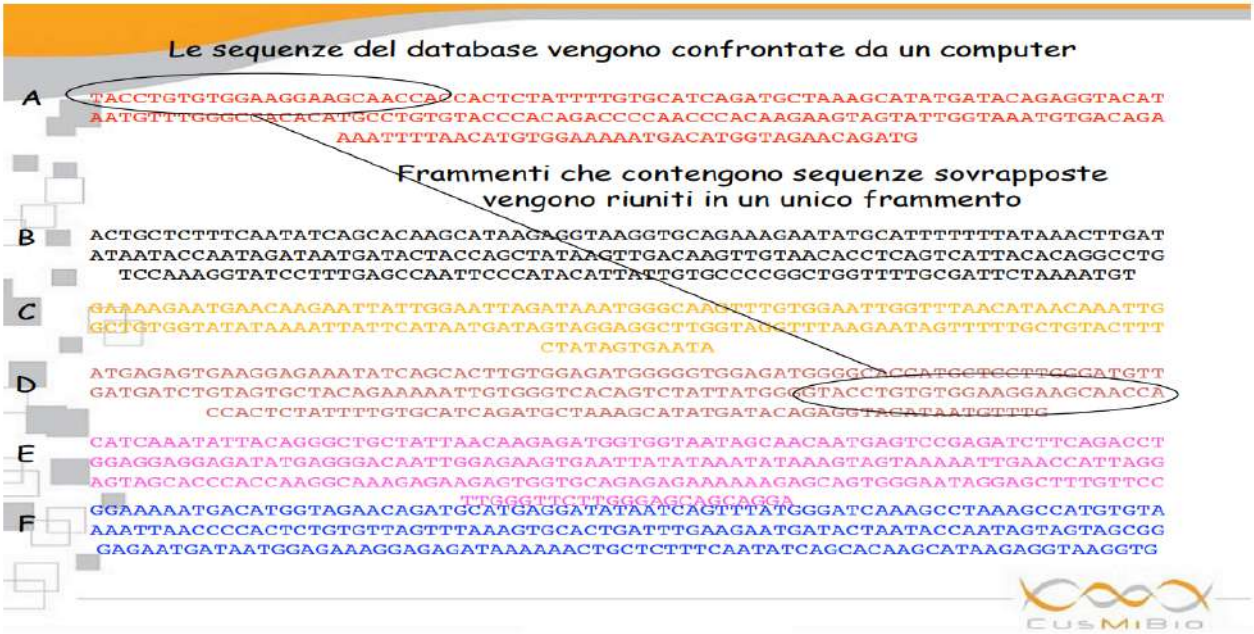
1. Il DNA umano totale è stato frammentato meccanicamente (spinto attraverso l'ago di una siringa), in frammenti di circa 500 bp. Per ogni regione del DNA si formano molti frammenti sovrapposti, per cui ogni sequenza viene sequenziata numerose volte;



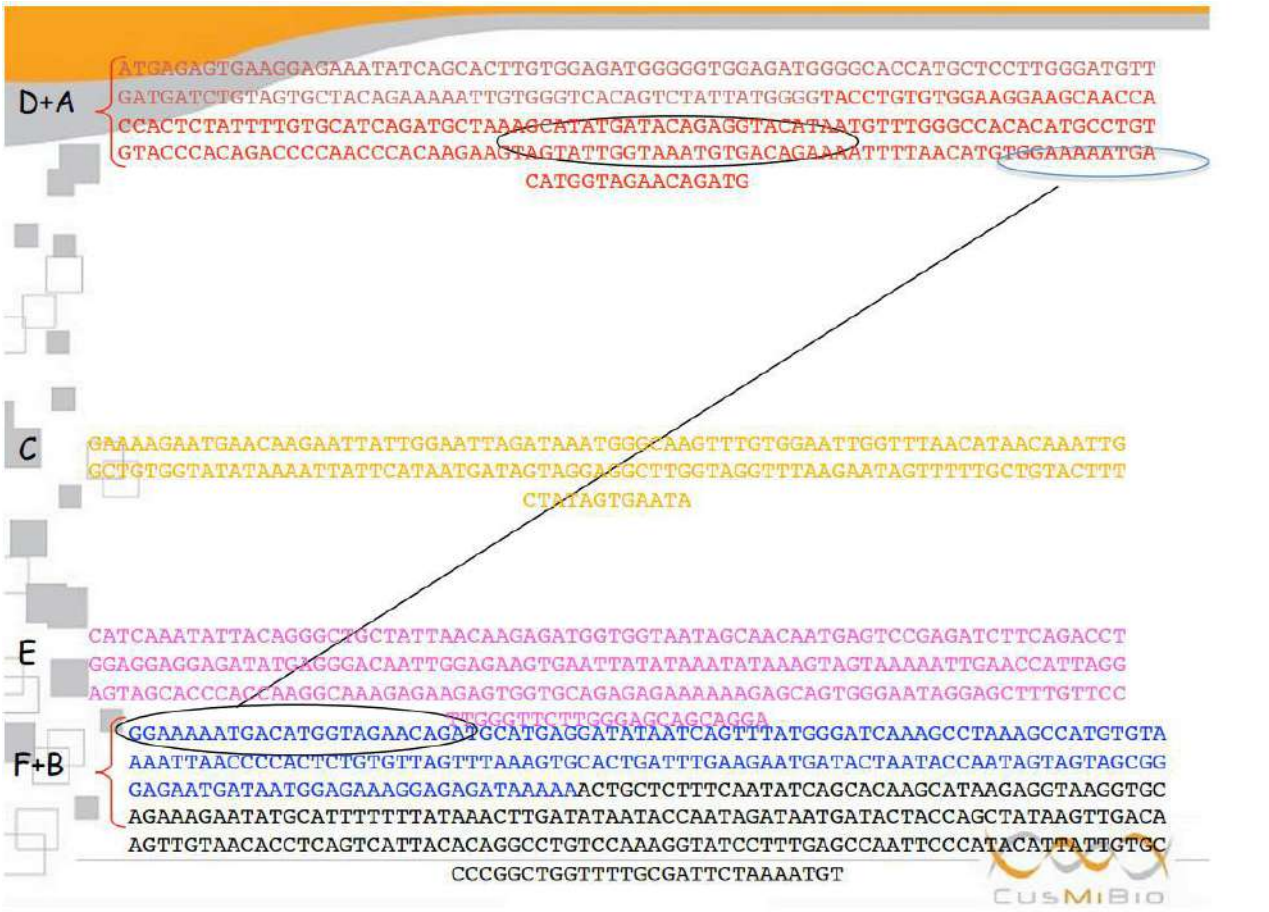
2. Ciascun frammento viene clonato in un vettore plasmidico e sequenziato automaticamente. Si ottengono così milioni di sequenze parzialmente sovrapposte;



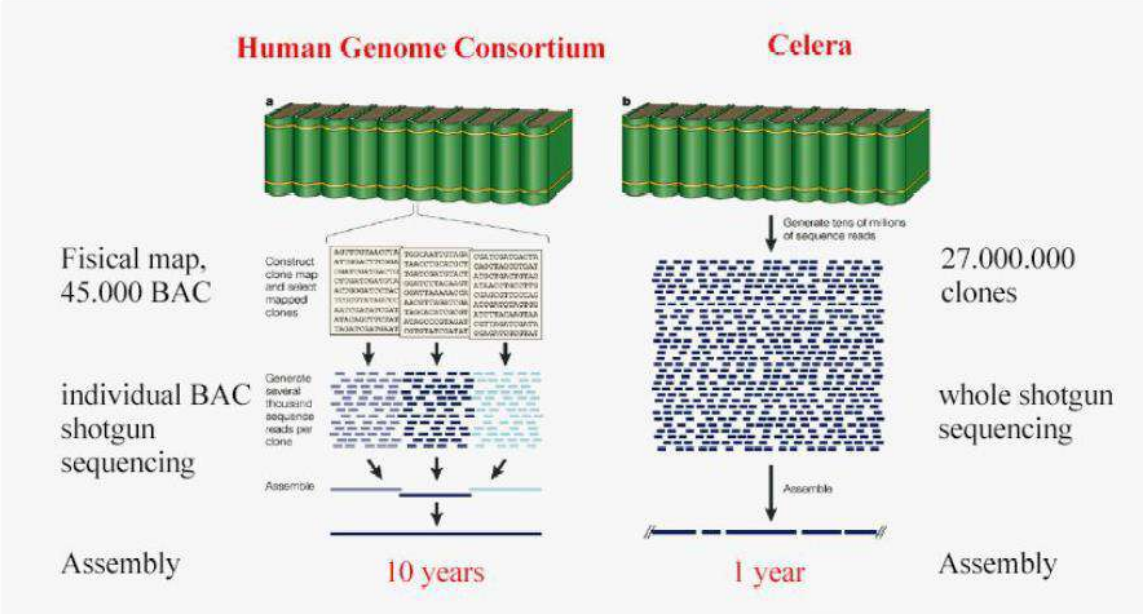
3. Le sequenze del database vengono confrontate da un computer, che riassembla in un'unica porzione frammenti che contengono sequenze sovrapposte.



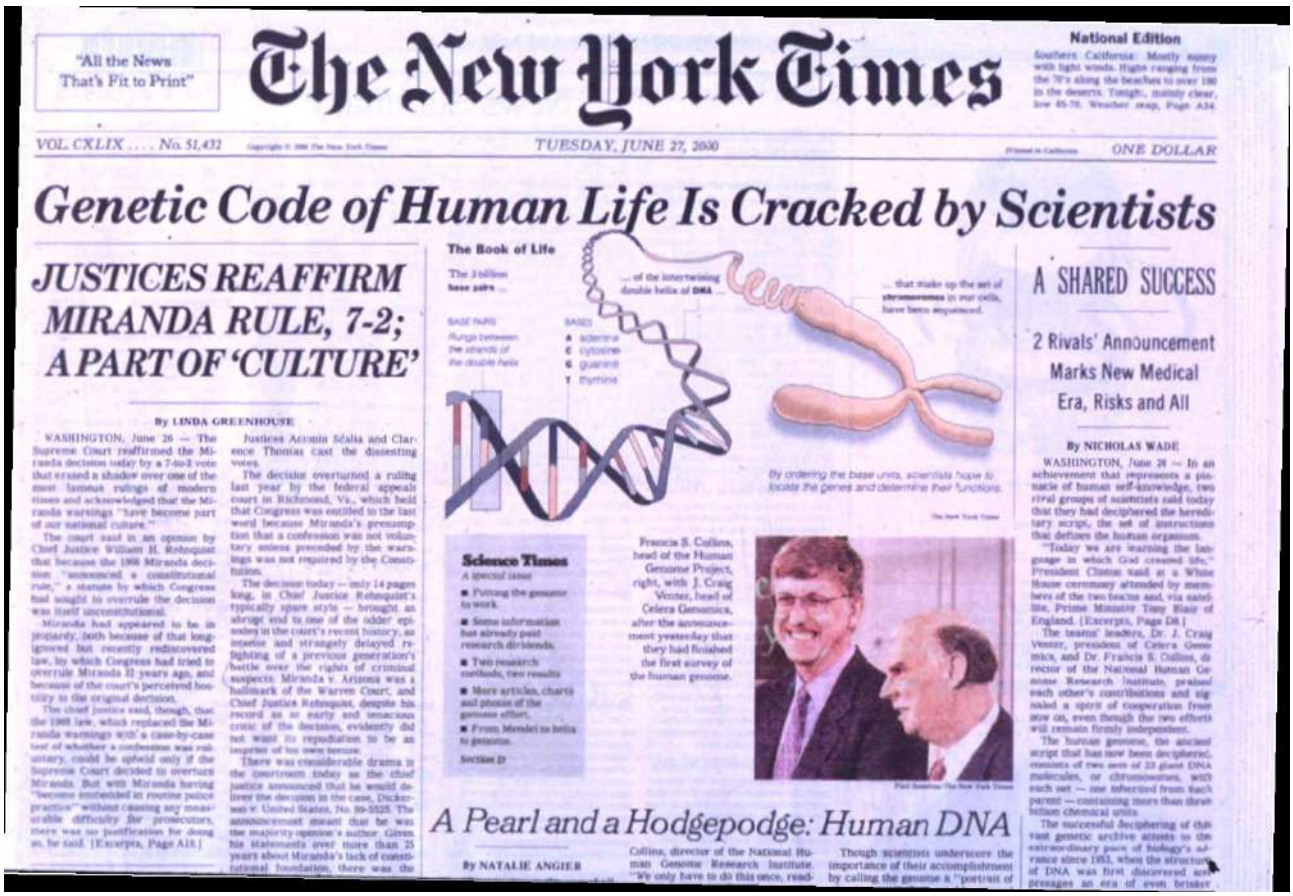
D+A
ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA
CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGA
CATGGTAGAACAGATG
B
ACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGCAGAAAGAATATGCATTTTTTTATAAACTTGAT
ATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACAAGTTGTAACACCTCAGTCATTACACAGGCCTG
TCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGCCCCGGCTGGTTTTGCGATTCTAAAATGT
C
GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTG
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTT
CTATAGTGAATA
E
CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC
TTGGGTTCTTGGGAGCAGCAGGA
F
GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGG
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTG


D+A
ATGAGAGTGAAGGAGAAATATCAGCACTTGTGGAGATGGGGGTGGAGATGGGGCACCATGCTCCTTGGGATGTT
GATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCA
CCACTCTATTTTGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGT
GTACCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTTAACATGTGGAAAAATGA
CATGGTAGAACAGATG
C
GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTG
GCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTT
CTATAGTGAATA
E
CATCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTAATAGCAACAATGAGTCCGAGATCTTCAGACCT
GGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGG
AGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCC
TTGGGTTCTTGGGAGCAGCAGGA
F+B
GGAAAAATGACATGGTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTA
AAATTAACCCCACTCTGTGTTAGTTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGTAGCGG
GAGAATGATAATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAGCATAAGAGGTAAGGTGC
AGAAAGAATATGCATTTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATAAGTTGACA
AGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCAATTCCCATACATTATTGTGC
CCCGGCTGGTTTTGCGATTCTAAAATGT

Ricapitolando, l'HGC si servì di circa 45 000 BAC e mappe fisiche. Il suo procedimento consistette nella costruzione di una mappa dei cloni e nella selezione dei cloni mappati, seguite dalla generazione di diverse migliaia di sequenze di ciascun clone, per poi terminare con l'assemblaggio dei frammenti. A Celera servirono 27 milioni di cloni e la generazione di decine di milioni di sequenze, che vennero successivamente assemblate.



Il 27 giugno 2000 arrivò l'annuncio: il genoma umano era stato sequenziato. Il 15 febbraio l'HGC pubblicò un articolo su "Nature", il giorno successivo Celera fece lo stesso su "Science", dove entrambi riportavano le scoperte avvenute.



"All the News That's Fit to Print"

The New York Times

National Edition

VOL. CXLIX . . . . No. 51,432 TUESDAY, JUNE 27, 2000 ONE DOLLAR

Genetic Code of Human Life Is Cracked by Scientists

JUSTICES REAFFIRM MIRANDA RULE, 7-2; A PART OF 'CULTURE'

By LINDA GREENHOUSE

WASHINGTON, June 26 — The Supreme Court reaffirmed the Miranda decision today by a 7-to-2 vote that erased a shadow over one of the most famous rulings of modern times and acknowledged that the Miranda warnings "have become part of our national culture."

The court said in an opinion by Chief Justice William H. Rehnquist that because the 1966 Miranda decision "announced a constitutional rule," a statute by which Congress had sought to overrule the decision was itself unconstitutional.

Miranda had appeared to be in jeopardy, both because of that long-ignored but recently rediscovered law, by which Congress had tried to overrule Miranda 33 years ago, and because of the court's perceived hostility to the original decision.

The chief justice said, though, that the 1968 law, which replaced the Miranda warnings with a case-by-case test of whether a confession was voluntary, could be upheld only if the Supreme Court decided to overturn Miranda. But with Miranda having "become embedded in routine police practice" without causing any measurable difficulty for prosecutors, there was no justification for doing so, he said. [Excerpts, Page A18.]

Justices Antonin Scalia and Clarence Thomas cast the dissenting votes.

The decision overturned a ruling last year by the federal appeals court in Richmond, Va., which held that Congress was entitled to the last word because Miranda's presumption that a confession was not voluntary unless preceded by the warnings was not required by the Constitution.

The decision today — only 14 pages long, in Chief Justice Rehnquist's typically spare style — brought an abrupt end to one of the odder episodes in the court's recent history, an intense and strangely delayed rerighting of a previous generation's battle over the rights of criminal suspects. Miranda v. Arizona was a hallmark of the Warren Court, and Chief Justice Rehnquist, despite his record as an early and tenacious critic of the decision, evidently did not want its repudiation to be an imprint of his own tenure.

There was considerable drama in the courtroom today as the chief justice announced that he would deliver the decision in the case, Dickerson v. United States, No. 99-5525. The announcement meant that he was the majority opinion's author. Given his statements over more than 25 years about Miranda's lack of constitutional foundation, there was the

The Book of Life

The 3 billion base pairs ... of the intertwining double helix of DNA ... that make up the set of chromosomes in our cells, have been sequenced.

BASE PAIRS Rungs between the strands of the double helix

BASES A adenine C cytosine G guanine T thymine

By ordering the base units, scientists hope to locate the genes and determine their functions.

Science Times

A special issue

- Putting the genome to work
- Some information has already paid research dividends.
- Two research methods, two results
- More articles, charts and photos of the genome effort.
- From Mendel to beta to genome.

SECTION D

Francis S. Collins, head of the Human Genome Project, right, with J. Craig Venter, head of Celera Genomics, after the announcement yesterday that they had finished the first survey of the human genome.

A SHARED SUCCESS

2 Rivals' Announcement Marks New Medical Era, Risks and All

By NICHOLAS WADE

WASHINGTON, June 26 — In an achievement that represents a pinnacle of human self-knowledge, two rival groups of scientists said today that they had deciphered the hereditary script, the set of instructions that defines the human organism.

"Today we are learning the language in which God created life," President Clinton said at a White House ceremony attended by members of the two teams and, via satellite, Prime Minister Tony Blair of England. [Excerpts, Page D8.]

The teams' leaders, Dr. J. Craig Venter, president of Celera Genomics, and Dr. Francis S. Collins, director of the National Human Genome Research Institute, praised each other's contributions and signaled a spirit of cooperation that now on, even though the two efforts will remain firmly independent.

The human genome, the ancient script that has now been deciphered, consists of two sets of 23 giant DNA molecules, or chromosomes, with each set — one inherited from each parent — containing more than three billion chemical units.

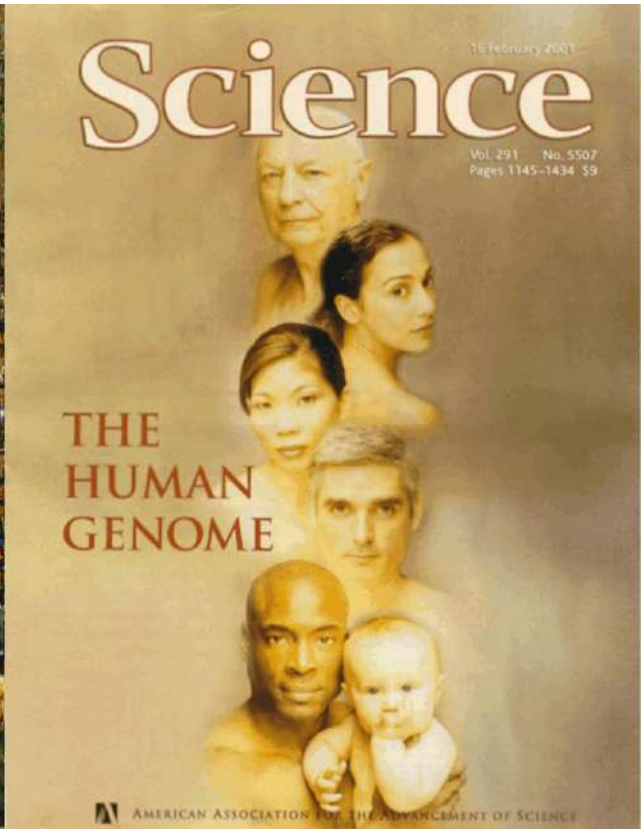
The successful deciphering of this vast genetic archive attests to the extraordinary pace of biology's advance since 1953, when the structure of DNA was first discovered and presages an era of even bolder

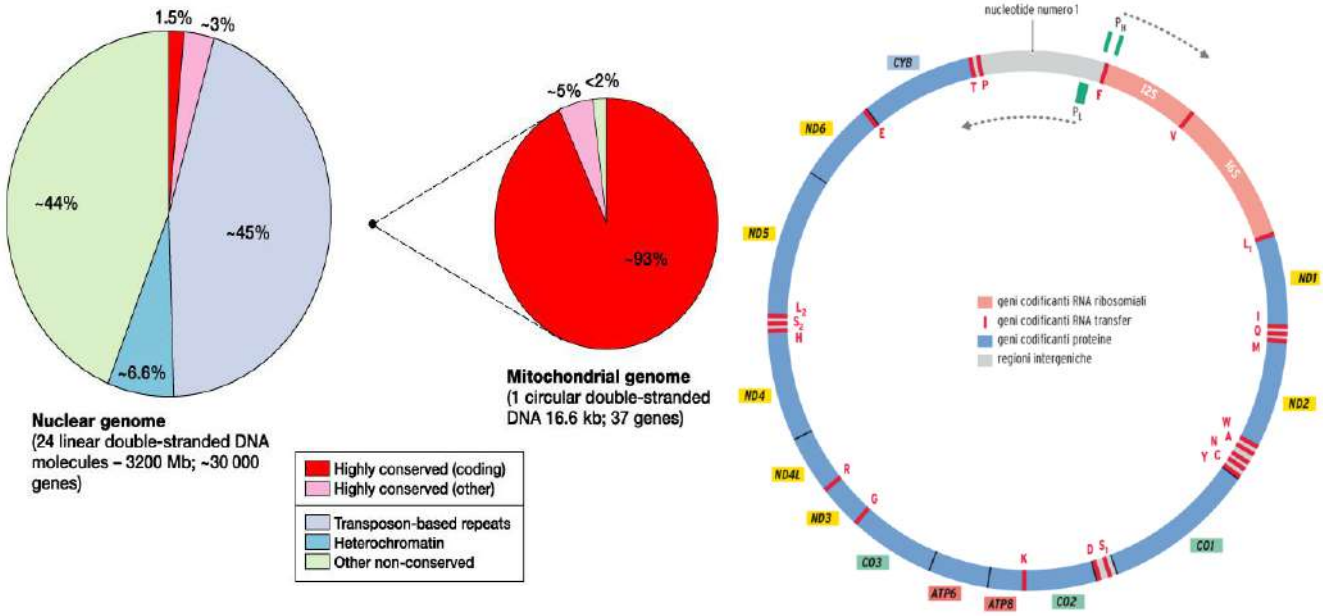
A Pearl and a Hodgepodge: Human DNA

By NATALIE ANGIER

Collins, director of the National Human Genome Research Institute. "We only have to do this once, read-

Though scientists underscore the importance of their accomplishment by calling the genome a "portrait of





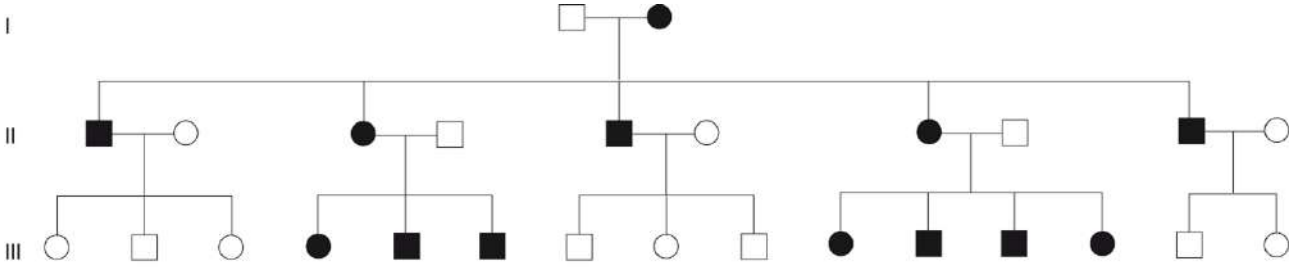
Autonomia limitata del genoma mitocondriale

	Codificati dal genoma mitocondriale	Codificati dal genoma nucleare
Componenti del sistema di fosforilazione ossidativa	13 subunità	> 80 subunità
I NADH deidrogenasi	7 subunità	> 41 subunità
II Succinato CoQ reduttasi	0 subunità	4 subunità
III Complesso citocromo b–c1	1 subunità	10 subunità
IV Complesso citocromo c ossidasi	3 subunità	10 subunità
V Complesso ATP sintasi	2 subunità	14 subunità
Componenti dell'apparato per la sintesi proteica	24	~ 80
Componenti dei tRNA	22 tRNA	Nessuno
Componenti degli rRNA	2 rRNA	Nessuno
Proteine ribosomiali	Nessuna	~ 80
Altre proteine mitocondriali	Nessuna	Tutte, ad es. DNA polimerasi e RNA polimerasi mitocondriali, più numerosi altri enzimi, proteine strutturali e di trasporto, ecc.

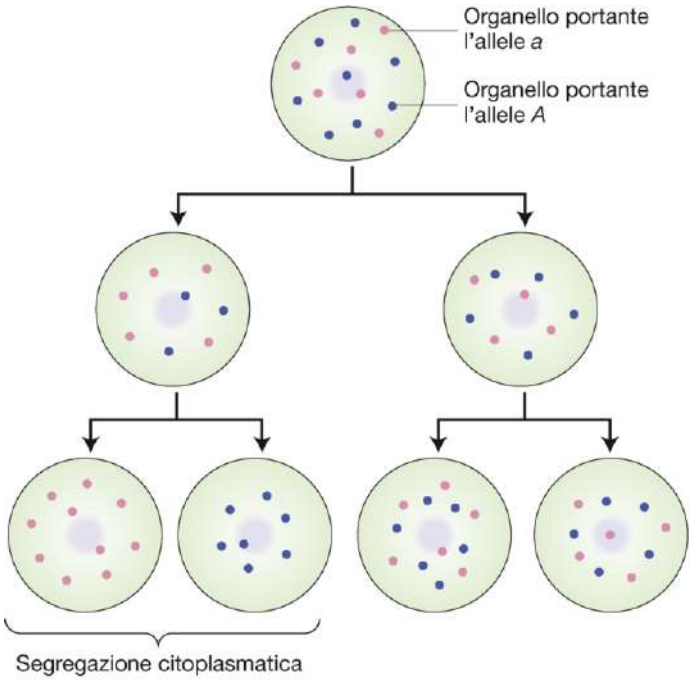
Non bisogna dimenticare che oltre al DNA nucleare esiste anche il DNA mitocondriale, composto da circa 16 kb, codificante per 37 geni, di cui due codificano per gli rRNA, 22 per i tRNA e 13 per polipeptidi. Il 93% del DNA mitocondriale è codificante. A questo organello è collegato il fenomeno della poliplasmia, ovvero sono presenti più copie del suo DNA all'interno del mitocondrio stesso. Gode di una limitata autonomia, poiché non è in grado di sintetizzare tutte le proteine necessarie per il corretto funzionamento.Possiede un proprio codice genetico, "privato". Viene ereditato secondo una discendenza matrilineare.

AAA	Lys	CAA	Gln	GAA	Glu	UAA	STOP
AAG		CAG		GAG		UAG	
AAC	Asn	CAC	His	GAC	Asp	UAC	Tyr
AAU		CAU		GAU		UAU	
ACA	Thr	CCA	Pro	GCA	Ala	UCA	Ser
ACG		CCG		GCG		UCG	
ACC		CCC		GCC		UCC	
ACU		CCU		GCU		UCU	
STOP AGA	Arg	CGA	Arg	GGA	Gly	Trp UGA	STOP
AGG		CGG		GGG		UGG	Trp
AGC	Ser	CGC		GGC		UGC	Cys
AGU		CGU		GGU		UGU	
Met AUA	Ile	CUA	Leu	GUA	Val	UUA	Leu
AUG	Met	CUG		GUG		UUG	
AUC	Ile	CUC		GUC		UUC	Phe
AUU		CUU		GUU		UUU	

I quadrati e i cerchi colorati indicano malattie correlate al DNA mitocondriale.

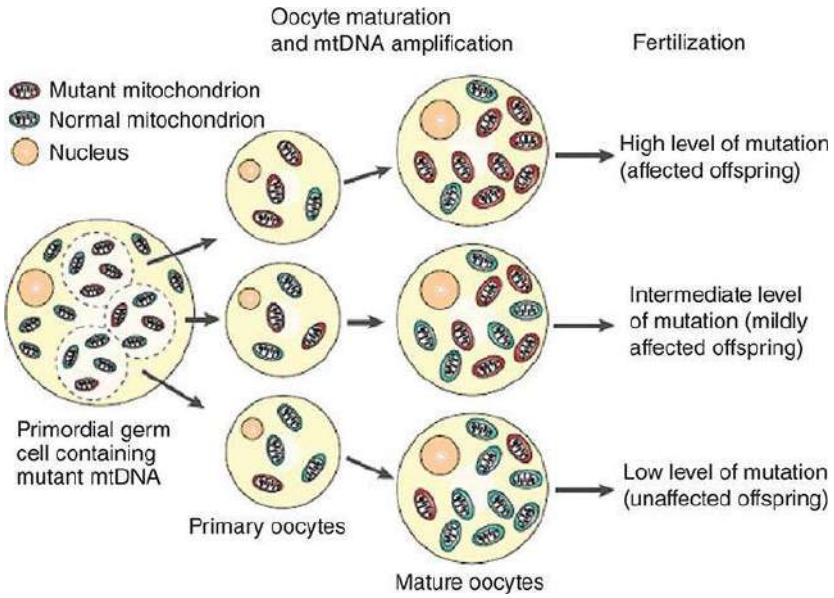


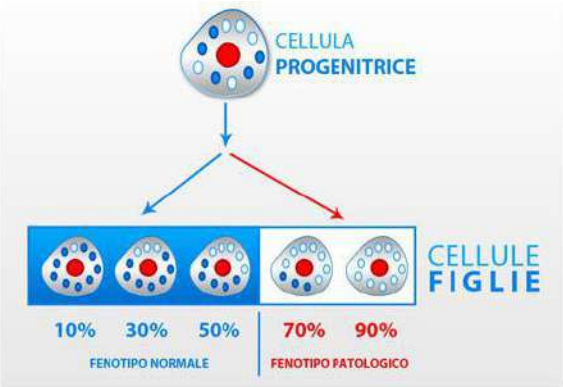
E' soggetto all'eteroplasmia, ovvero sono presenti più mitocondri all'interno della stessa cellula. Segrega casualmente nel citoplasma al momento della divisione cellulare.



I vari mitocondri si distribuiscono al momento della divisione secondo un effetto "collo di bottiglia".

Quando il DNA mitocondriale eteroplasmico viene trasmesso da una madre ai figli, la percentuale di una determinata sequenza che segrega negli oociti cambia in modo stocastico attraverso un fenomeno detto di *bottleneck* (letteralmente, "collo di bottiglia"). In pratica, se in una determinata posizione del genoma mitocondriale è presente una mutazione e il DNA mutato è contenuto nelle cellule somatiche della madre ad esempio all'80 per cento, nella linea germinale non si mantiene la stessa percentuale. Un oocita potrebbe contenere il DNA mutato nel 10 per cento dei suoi mitocondri, un altro al 100 per cento, uno non averne affatto e così via.

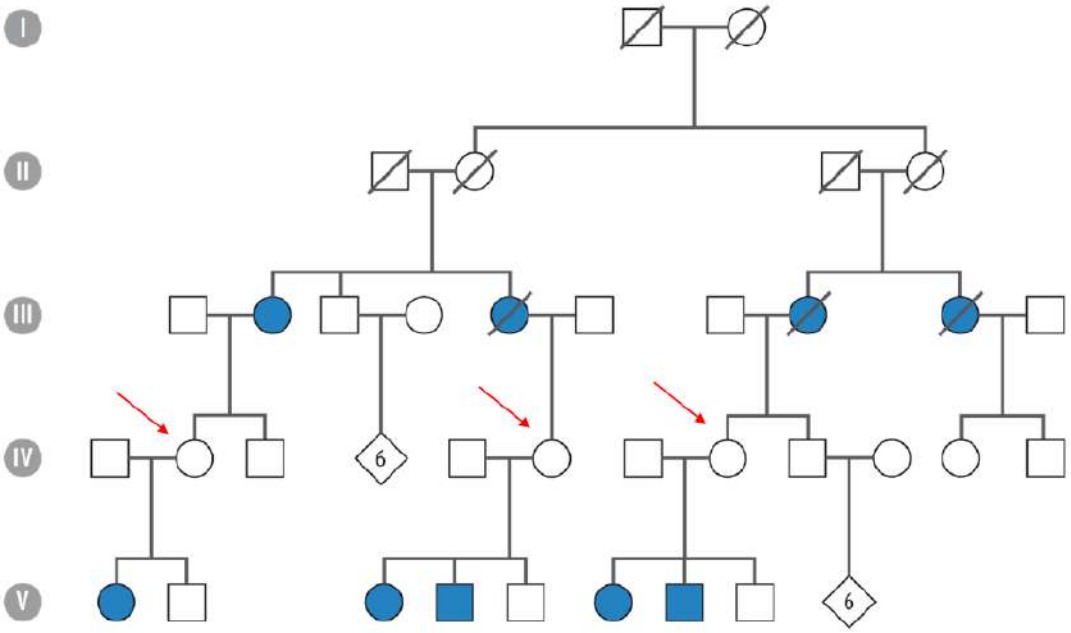




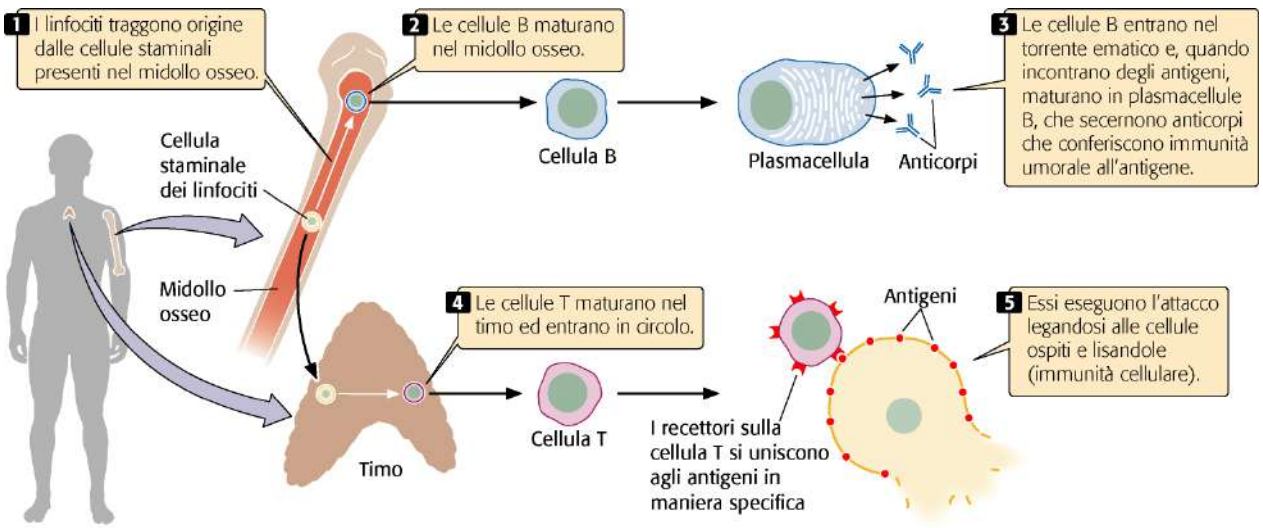
Esiste anche il cosiddetto effetto "soglia", che si verifica quando alcuni mitocondri sono portatori di mutazioni nel proprio DNA. Se il numero di mitocondri mutanti ereditati dalla madre è sotto una soglia determinata (ad esempio, del 70%), l'individuo in esame non presenta alcun fenotipo patologico, mentre se la percentuale supera il 70% il fenotipo sarà malato.

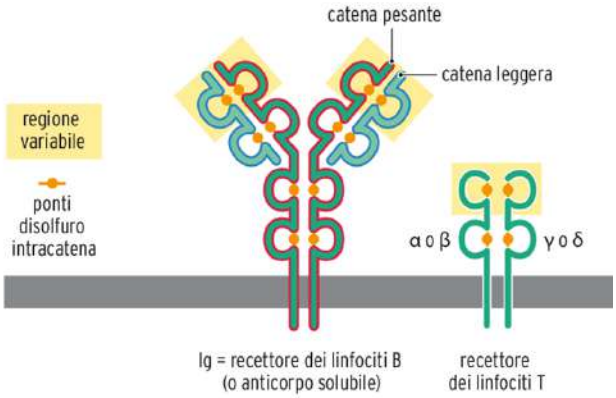
5. Ereditarietà matrilineare **8. Effetto collo di bottiglia**

9. Effetto soglia



Classi di proteine del sistema immunitario





numero aploide di geni umani	3*	4*
numero delle diverse proteine	molti milioni in una persona	milioni in una persona
fonte della variazione genetica	ricombinazione somatica e ipermutazione	ricombinazione somatica

I linfociti traggono origine dalle cellule staminali presenti nel midollo osseo. Le cellule B maturano nel midollo osseo. Le cellule B entrano nel torrente ematico e, quando incontrano antigeni, maturano in plasmacellule B, che secernono anticorpi e conferiscono immunità umorale all'antigene. Le cellule T, invece, maturano nel timo ed entrano in circolo. I recettori su queste cellule si uniscono agli antigeni in maniera specifica, così attaccano la cellula ospite e la lisano (immunità cellulare).

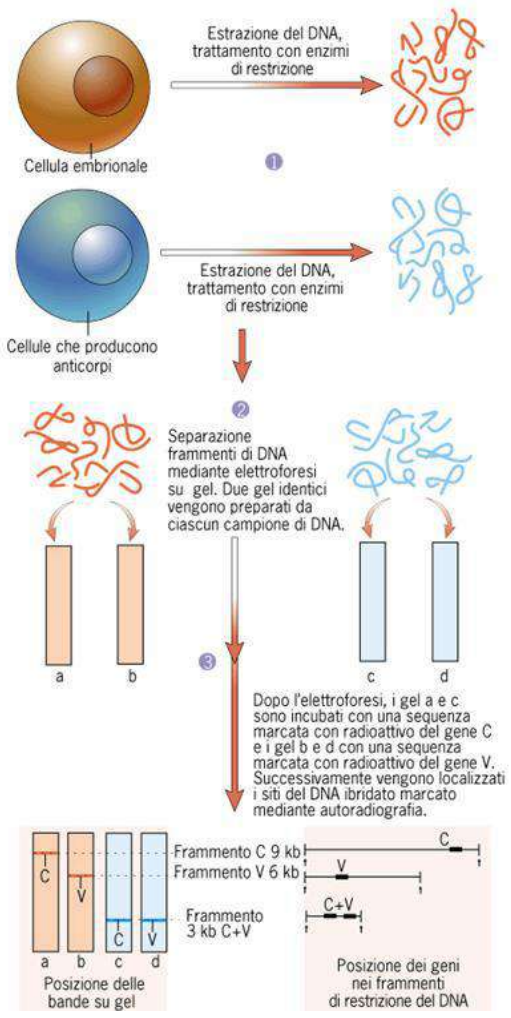
I recettori delle cellule B vengono detti anticorpi solubili, o anche immunoglobuline (Ig). Ciascuna di esse è composta da due catene pesanti e da due catene leggere, più una regione variabile e ponti disolfuro intracatena. Sono codificate da tre geni su ciascun cromosoma umano (i recettori delle cellule T sono codificate da 4 per cromosoma) e si formano molti milioni di proteine diverse per ogni persona.

Esistono cinque classi di Ig: IgA, IgD, IgE, IgG, IgM.

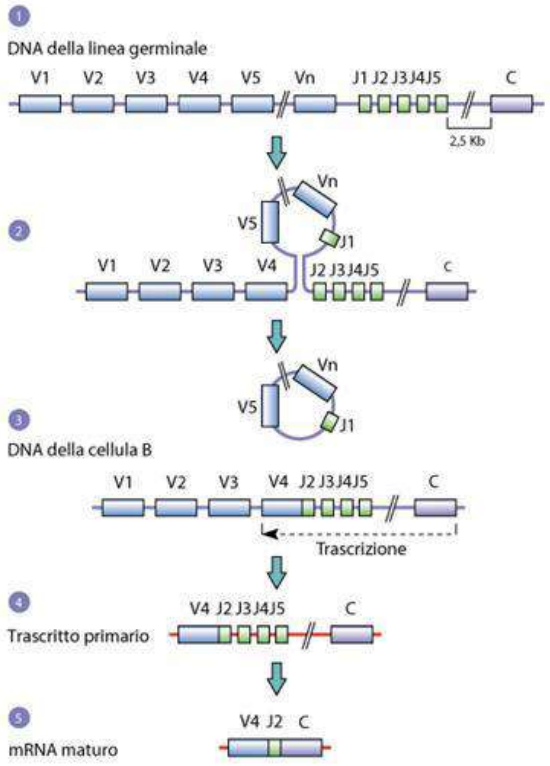
TABLE 18.3

Antibody Classes (Part 1)

CLASS	GENERAL STRUCTURE		LOCATION	FUNCTION
IgG	Monomer		Free in blood plasma; about 80 percent of circulating antibodies	Most abundant antibody in primary and secondary immune responses; crosses placenta and provides passive immunizationto fetus
IgM	Pentamer		Surface of B cell; free in blood plasma	Antigen receptor on B cell membrane; first class of antibodies released by B cells during primary response
IgD	Monomer		Surface of B cell	Cell surface receptor of mature B cell; important in B cell activation
IgA	Dimer		Saliva, tears, milk, and other body secretions	Protects mucosal surfaces; prevents attachment of pathogens to epithelial cells
IgE	Monomer		Secreted by plasma cells in skin and tissues lining gastrointestinal and respiratory tracts	When bound to antigens, binds to mast cells and basophils to trigger release of histamine that contributes to inflammation and some allergic responses

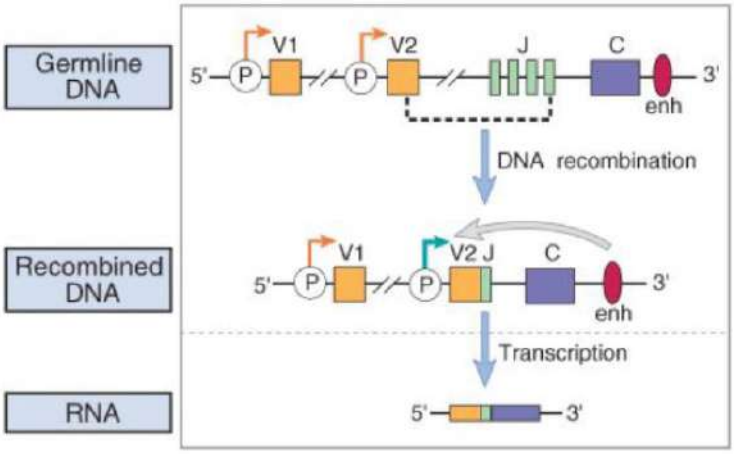


Tonegawa e Hozumi, nel 1976, dimostrarono che, nelle cellule B, i geni Ig sono riarrangiati. Estrassero e trattarono con enzimi di restrizione il DNA proveniente da una cellula embrionale e da una cellula che produceva anticorpi. Si operò la separazione dei frammenti mediante elettroforesi su gel: due gel identici vennero preparati per ciascun campione di DNA. Dopo l'elettroforesi due geni vennero incubati con una sonda marcata con un radioattivo del gene C (costante) e gli altri due geni vennero incubati con una sonda marcata con un radioattivo del gene V (variabile). Successivamente vennero localizzati i siti del DNA ibridato marcato mediante autoradiografia. Il frammento C era lungo 9 kb, quello V 6 kb (entrambi provenienti dalla cellula embrionale), mentre il risultato dei frammenti della cellula B fu un sorprendente 3 kb, formato dalla somma di due geni C + V. Era avvenuto a monte un evento di ricombinazione somatica. La scoperta di questo meccanismo genetico di diversità anticorpale valse a Tonegawa il Nobel nel 1987.

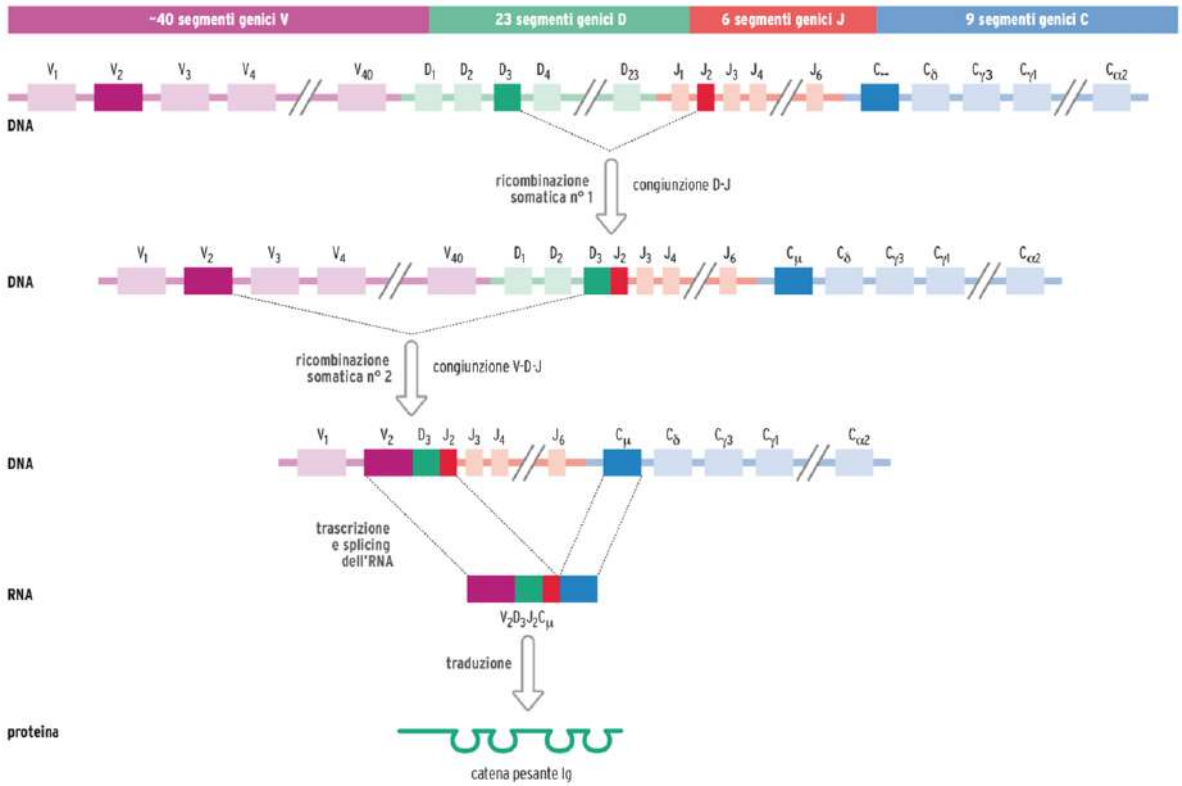


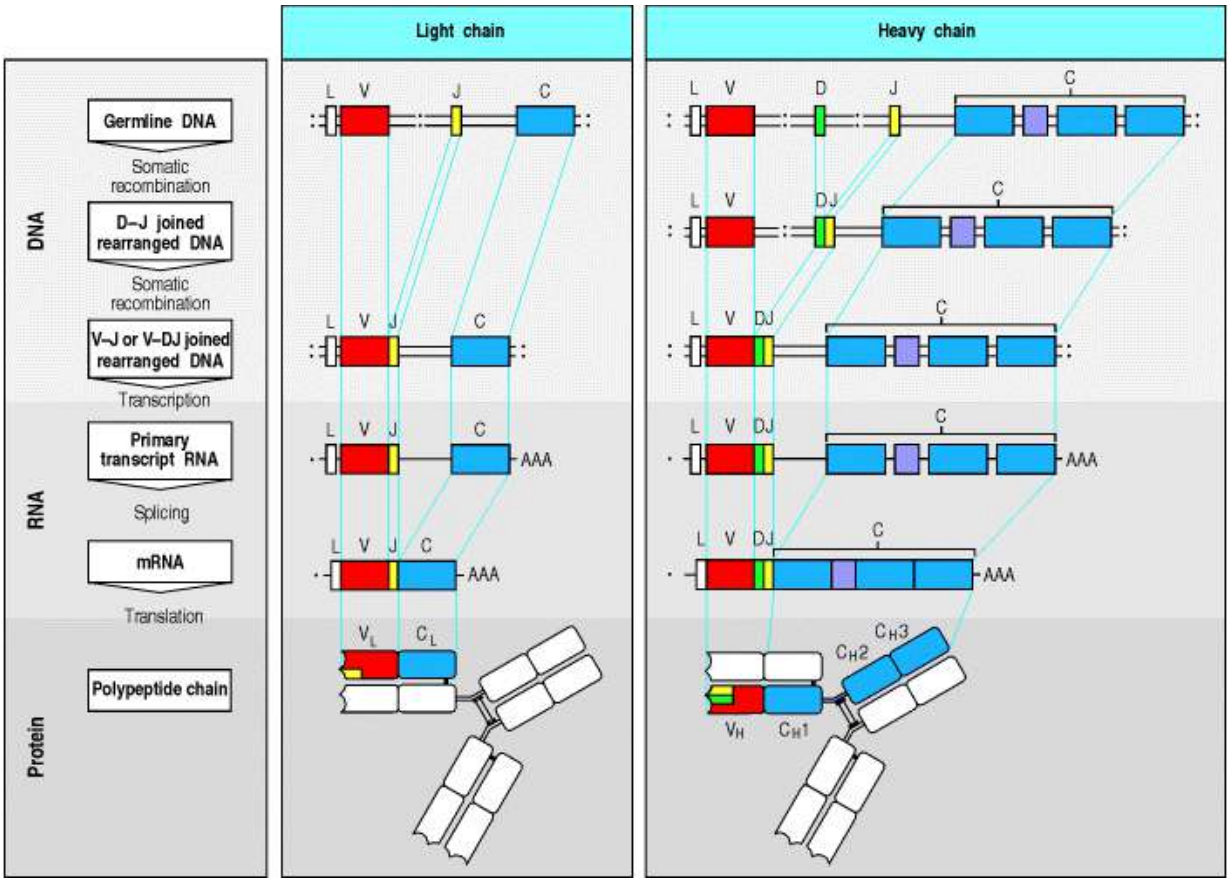
Il fenomeno che coinvolge il locus IGK (sul cromosoma 2), che codifica per la catena leggera di un anticorpo è, appunto, di ricombinazione somatica. Il locus è composto, in sequenza da circa 76 geni V (della regione variabile) e 5 geni J (joining region). Questi geni sono separati nel DNA genomico, cioè nella configurazione della linea germinale e compongono famiglie multigeniche. In primo luogo il DNA è riarrangiato; questo lo rende possibile alla congiunzione di 1 segmento genico V e 1 segmento genico J; le sequenze intermedie sono poi delete. L'RNA pre-messaggero è copiato (trascrizione); questo include gli introni. Poi avviene lo splicing; l'eliminazione degli introni dall'RNA premessaggero, che diviene maturo, definito RNA messaggero. In seguito poi avviene la sintesi proteica (conosciuta come traduzione). Il processo

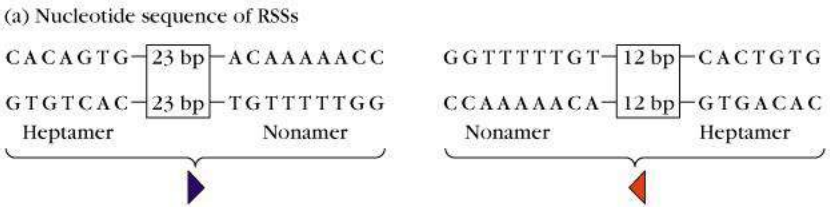
di selezione è casuale e al termine del riarrangiamento si formerà un singolo esone che codificherà per la regione variabile di un recettore per l'antigene. A valle di queste sequenze si trova il gene C, che codifica per la porzione costante, e ancora più a valle si incontra l'enhancer che promuove la trascrizione dell'intera sequenza dei geni.



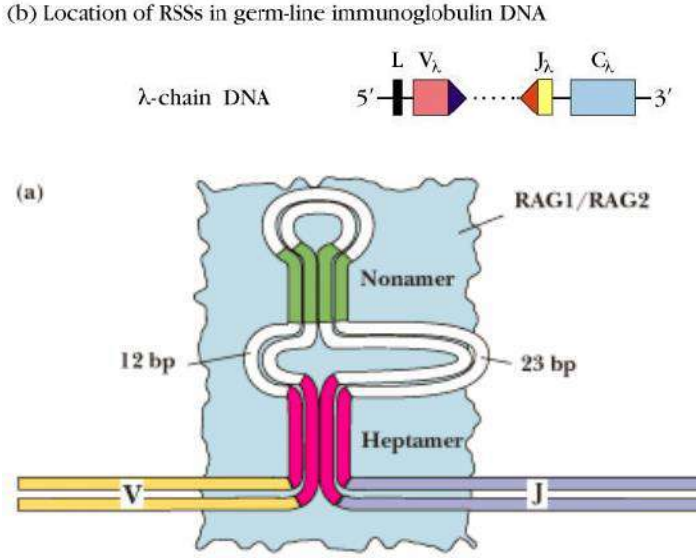
Nel caso del locus IGH (sul cromosoma 14), che codifica per la catena pesante di un anticorpo, si assiste ad un doppio evento di ricombinazione somatica. I segmenti genici che si possono combinare sono 40 V, 23 D, 6 J e 9 C, disposti in questo ordine lungo il DNA, leggendo da monte a valle la sequenza. Il primo riarrangiamento coinvolge un gene D e un gene J, che si congiungono, mentre la porzione intermedia compresa fra i due geni "selezionati" viene escissa sotto forma di "loop". La successiva ricombinazione somatica avviene tra il neonato blocco DJ e un gene V, cioè si parla di ricombinazione VDJ. Questo complesso, insieme a C, verrà trascritto, sottoposto a splicing e infine tradotto, a formare la catena pesante di un anticorpo.



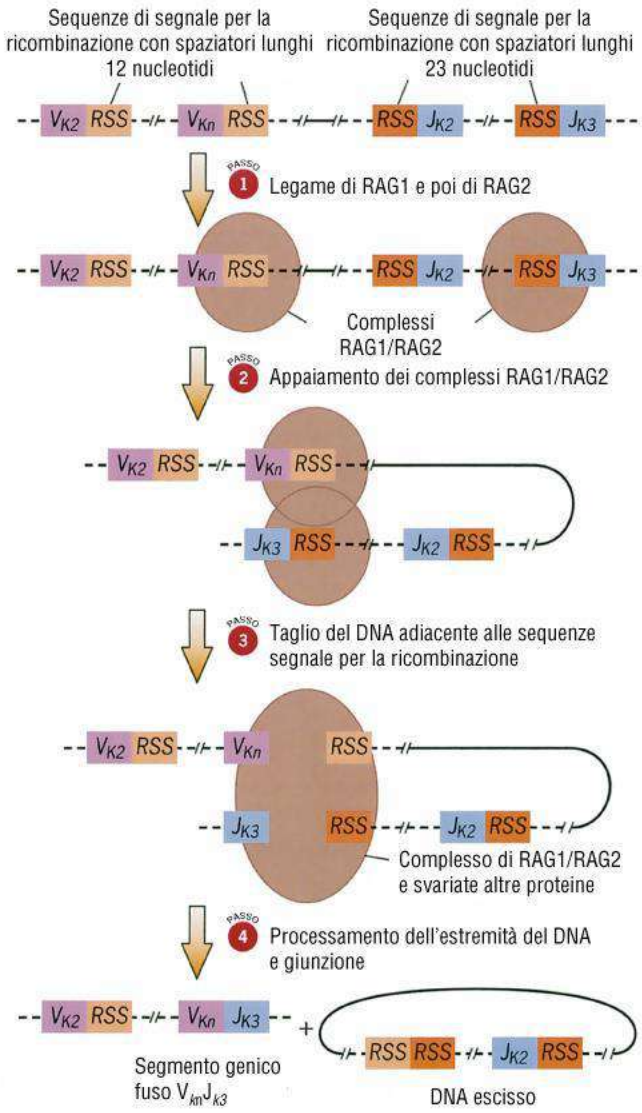




I meccanismi molecolari attraverso i quali si attua la ricombinazione V(D)J coinvolgono sequenze note con il nome di RSS (sequenze di segnale di ricombinazione), che includono una sequenza eptamerica e una sequenza nonamerica, separate da uno spaziatore. Ogni gene IGKV è seguito a valle (in posizione 3') da un RS costituito da un eptamero CACAGTG, e poi da uno spaziatore di 12 bp,e infine da un nonamero ACAAAAACC.

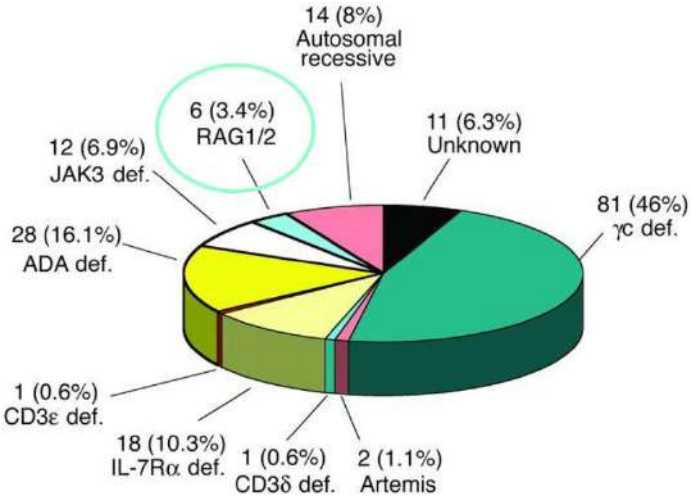


Ogni gene IGKJ è preceduto a monte (in posizione 5') da un RS costituito, tra il 5' e il 3', da un nonamero GGTTTTTGT, uno spaziatore di 23 bp e un eptamero CACTGTG. L'organizzazione di queste sequenze-segnale di ricombinazione (RSSs) è tale che la ricombinazione unisce sempre un segnale 12 a un segnale 23, regola12/23. I principali enzimi del riarrangiamento sono RAG1 e RAG1 e sono



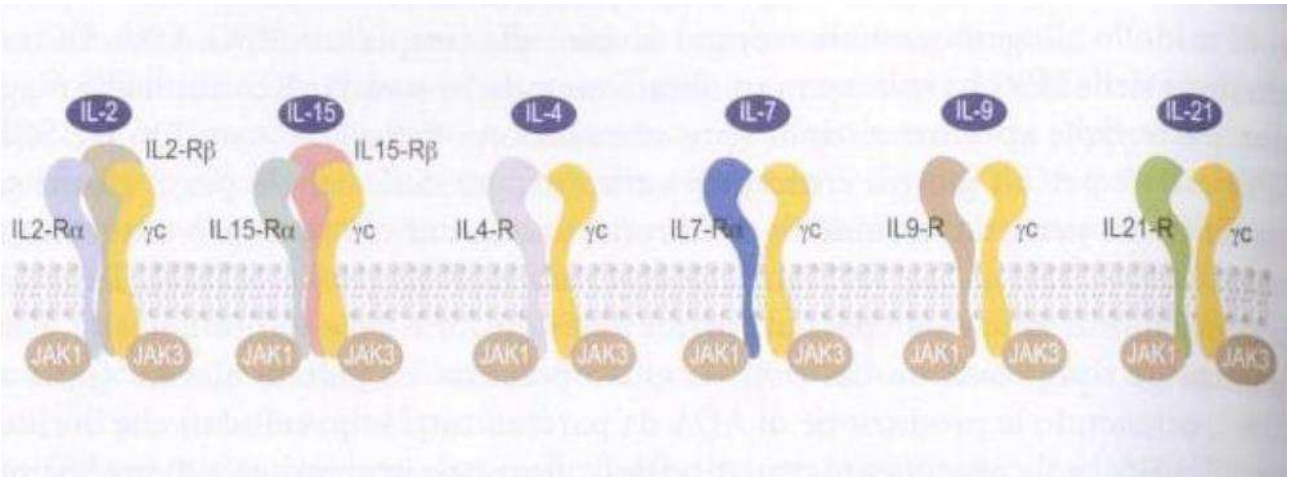
dette ricombinasi. Questi enzimi vengono espressi transitoriamente durante la maturazione dei T e dei B. RAG1 e, successivamente, RAG 2, si legano alle due sequenze RSS coinvolte. I due complessi si appaiano, unendo un segnale 12 a un segnale 23. Viene effettuato il taglio del DNA adiacente alle RSS. Avviene una fase di processamento di DNA alle sue estremità e infine si saldano le giunzioni. Così si ottiene un segmento genico fuso VJ e un anello di DNA escisso.

Diversi sono i fattori che contribuiscono alla diversità delle Ig: geni multipli per la regione variabile V, D, J, la libera combinazione tra le catene H e le catene L, la flessibilità del punto di giunzione tra V e J, tra D e J, tra V e D. Durante il riarrangiamento genico (VDJ) si possono verificare errori di saldatura e questo aumenta la variabilità della catena prodotta. Delle nucleasi possono eliminare alcuni nucleotidi e l'enzima TdT (Trasferasi deossinucleotide Terminale) ne può aggiungere nuovi (nucleotidi N). Inoltre quando RAG taglia la forcina del DNA e ripara il DNA interrotto può aggiungere sequenze di nucleotidi (nucleotidi P). Tutto ciò permette all'essere umano di sviluppare 10^13 tipi di Ig.



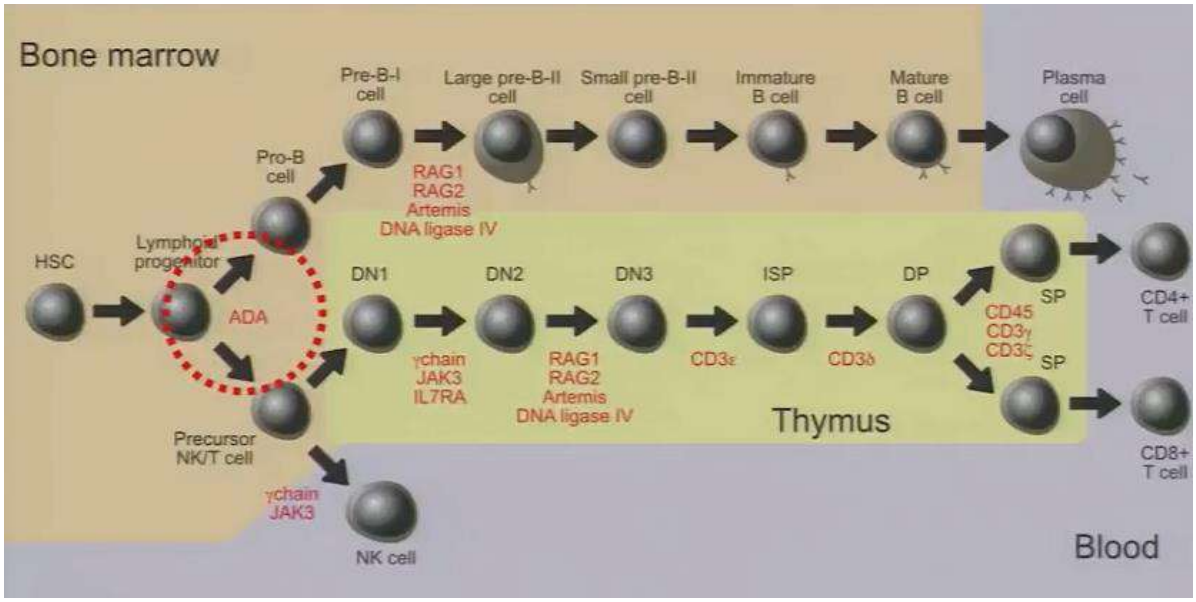
Il 3,4% delle mutazioni che provocano la SCID (Severe Combined Immunodeficiency Disease) sono a carico del complesso RAG1/2. La percentuale (46%) più sostanziosa dei casi è però riconducibile a un tipo di SCID legata al cromosoma X, che prende appunto il nome di X SCID. Questa condizione, recessiva, è dovuta a una mutazione della catena gammaC del recettore dell'interleuchina 2 (IL2RG). I pazienti affetti da questa patologia presentano un quadro di deficit immunitario a carico dei linfociti B e T; le aspettative di vita

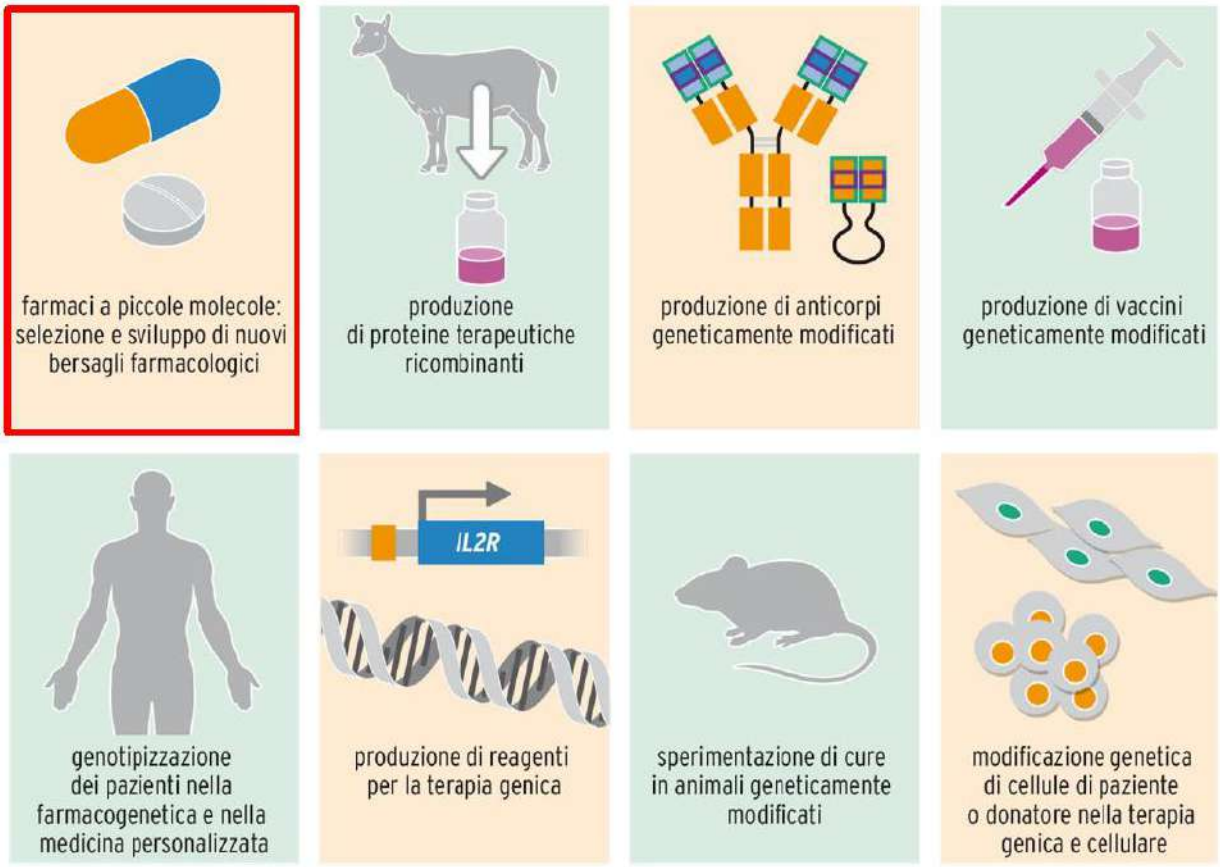
sono piuttosto scarse e, in genere, le complicanze derivano da infezioni opportunistiche, anche se la sopravvivenza dei pazienti sono piuttosto ridotte ed è possibile un approccio ex vivo.



Emblematico fu il caso di David Philip Vetter, nato nel settembre 1971 e morto nell'ottobre 1984, dopo aver vissuto tutta la propria esistenza in una bolla perché affetto da SCID.

Le mutazioni a carico del gene ADA costituiscono il 16,1% delle mutazioni che causano SCID, provocando ADA-SCID. L'ADA-SCID è una rara patologia che appartiene al gruppo delle immunodeficienze severe combinate (SCID), malattie in cui il sistema immunitario è gravemente compromesso, al punto che l'organismo è incapace di difendersi dagli agenti infettivi. Nella maggior parte dei casi, l'ADA-SCID si manifesta già a partire dai primi mesi di vita con infezioni ricorrenti, spesso sostenute da germi normalmente innocui per l'uomo e caratterizzate da un decorso particolarmente aggressivo. La malattia è causata dall'alterazione del gene ADA, che permette la produzione di un enzima chiamato adenosina deaminasi (ADA), importante per la maturazione e la funzionalità dei linfociti, cellule del sistema immunitario fondamentali per la difesa dell'organismo dalle infezioni. La malattia si trasmette come carattere autosomico recessivo: questo significa che se in una coppia entrambi i genitori sono portatori sani, a ogni gravidanza si ha un rischio del 25 per cento di generare figli affetti.

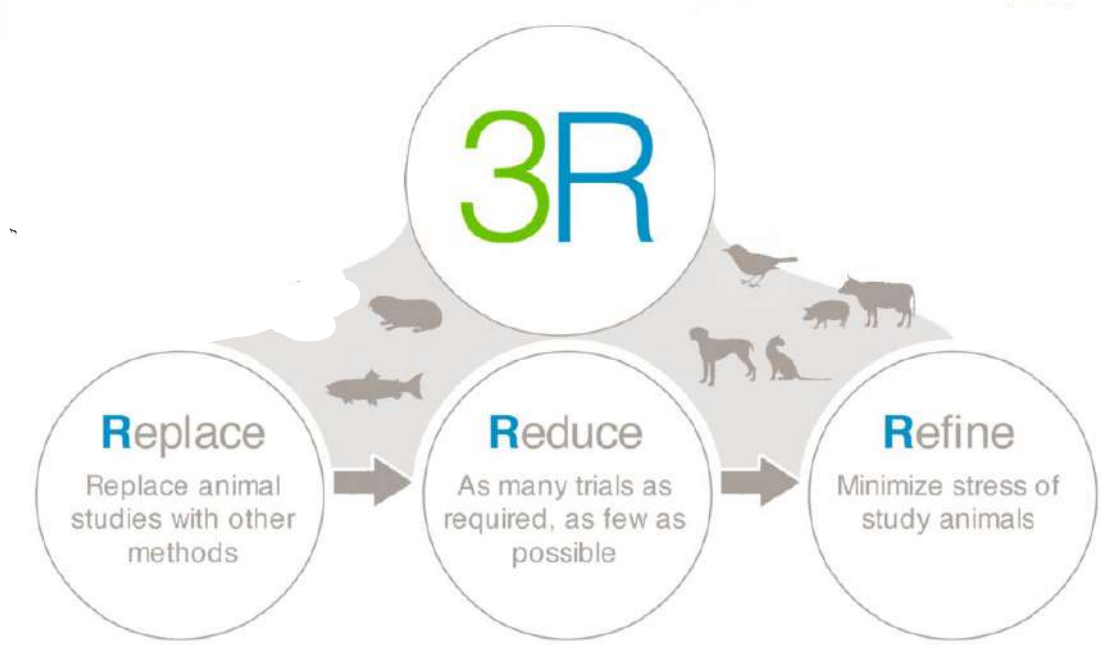




Nuovi approcci genetici sono stati sviluppati al trattamento delle malattie, ad esempio l'ideazione di farmaci a piccole molecole, connessa alla selezione e allo sviluppo di nuovi bersagli farmacologici. Gli stadi principali per lo sviluppo di un farmaco sono i seguenti:

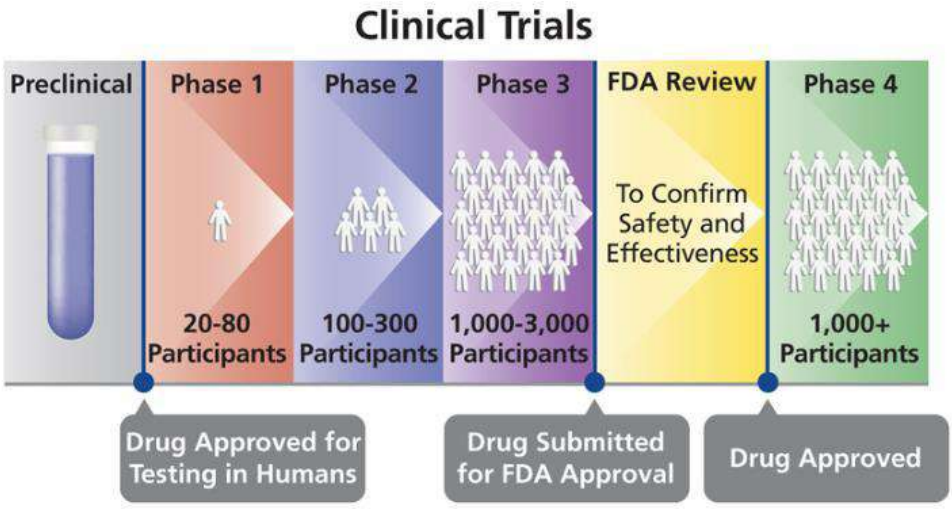
1. La prima fase è preclinica, incentrata su test di laboratorio su animali e su cellule in coltura per testare la tossicità, la farmacocinetica e l'efficacia del farmaco. In merito ai test di laboratorio sugli animali esiste una direttiva, la n.63 del 2010, dell'Unione Europea, che mira

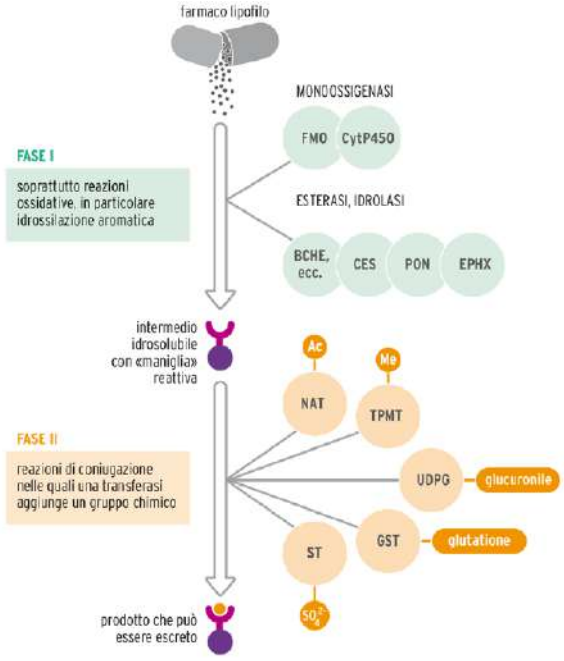
alla protezione degli animali utilizzati a fini scientifici, secondo il principio delle 3 R. La prima R sta per replace, un invito a sostituire gli studi condotti sugli animali con altre metodi, la seconda R sta per reduce, un suggerimento a compiere un numero sufficiente di prove perché siano valide, compiendone il meno possibile; l'ultima R sta per refine, riferito allo stress degli animali che va minimizzato. Seguono le sperimentazioni cliniche.



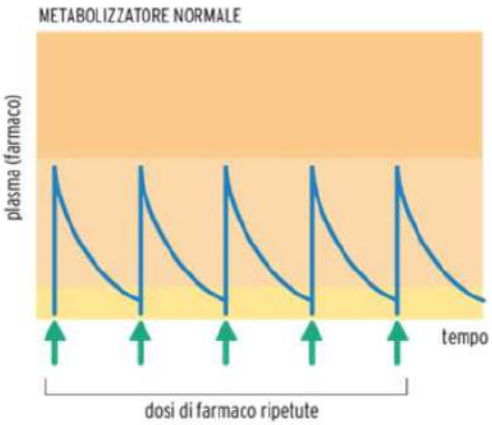
2. La fase 1 clinica coinvolge fino a 100 volontari sani, per controllare sicurezza generale, tollerabilità, farmacodinamica e cinetica;
3. La seconda fase clinica riguarda diverse centinaia di pazienti per assicurarsi della sicurezza e dell'efficacia del farmaco;
4. L'ultima fase avviene in migliaia di pazienti in sperimentazioni randomizzate controllate multicentriche, sempre volta a verificare sicurezza ed efficacia del farmaco.

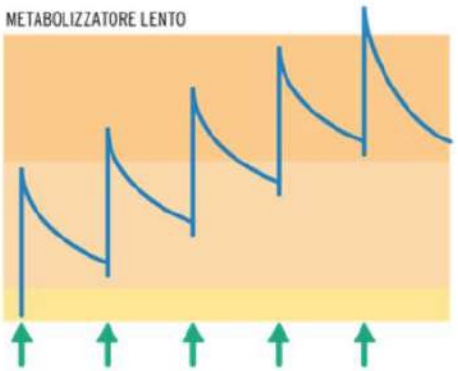
L'intero processo ha un costo che si aggira tra i 500 e i 700 milioni di dollari e ha una durata media di 15 anni. L'ultima approvazione arriva da parte della Food and Drug Administration, che permette la messa in commercio del farmaco.

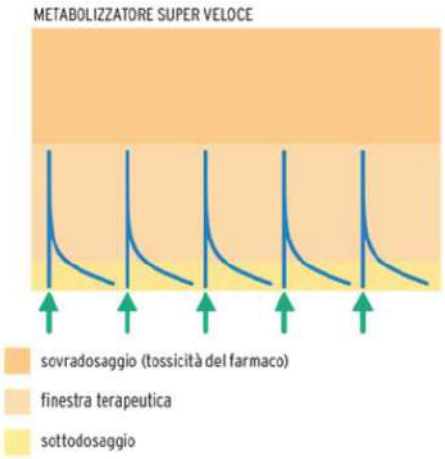




A questo tradizionale metodo di cura delle malattie si affianca la nascente frontiera della genotipizzazione dei pazienti nella farmacogenetica e nella medicina personalizzata. Questo si rende necessario anche per soprassedere ai problemi di metabolismo dei farmaci, che possono essere pericolosi e talvolta mortali in relazione al genotipo del paziente. Il metabolismo dei farmaci si compone di due fasi: nella prima il farmaco lipofilo è bersaglio soprattutto di reazioni ossidative, in particolare dell'idrossilazione aromatica. Si ottiene così un intermedio idrosolubile con una "maniglia" reattiva. La seconda fase consiste in reazioni di coniugazione nelle quali una transferasi aggiunge un gruppo chimico. Infine, si giunge ad un prodotto che può essere escreto. La velocità a cui avvengono questi passaggi è variabile: può essere normale, lento (in tal caso è facile provocare una condizione di sovradosaggio, che è collegata ad una tossicità da parte del farmaco) o veloce, che smaltisce più rapidamente il farmaco, dunque necessiterà di una somministrazione più frequente che, però, non può essere possibile oltre un dato numero di volte. Queste problematiche verranno presto superate dalla produzione di reagenti per la terapia genica o dalla modificazione genetica di cellule di paziente o donatore nella terapia genica e cellulare.

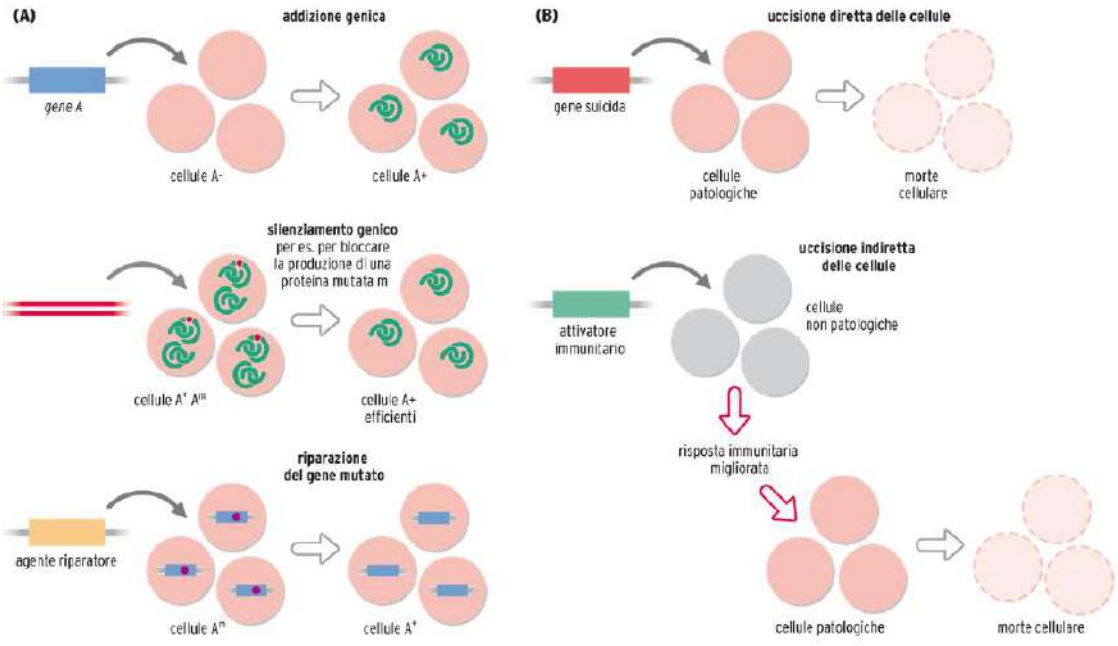




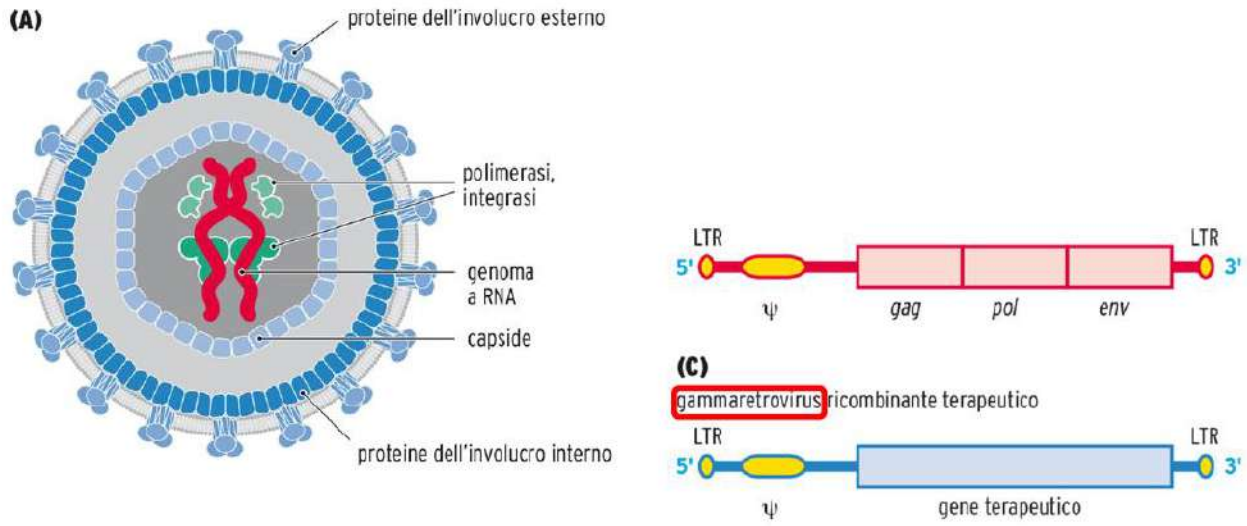


DANNO O TOSSICITÀ INDOTTI DA FARMACI	ESEMPI DI FARMACI ASSOCIATI	COMMENTI ED ESEMPI
Apnea (paralisi respiratoria)	sussametonio (succinilcolina)	questo farmaco agisce come miorilassante ad azione rapida ed è utilizzato prima di un intervento chirurgico. Di solito gli effetti del farmaco svaniscono rapidamente quando il farmaco viene metabolizzato dall'enzima butirrilcolinesterasi. I metabolizzatori lenti sono a rischio di apnea: dopo l'intervento chirurgico restano paralizzati e incapaci di respirare poiché non possono riacquistare la funzione muscolare abbastanza rapidamente e possono richiedere una ventilazione prolungata
Prolungamento dell'intervallo cardiaco QT	tioridazina, claritromicina, terfenadina	indotto da molti farmaci diversi. Associato alla tachicardia ventricolare polimorfica, o a torsioni di punta (Figura 1) che possono essere fatali
Tossicità ematologica	6-mercaptopurina, azatioprina	la tiopurina S-metiltransferasi (TPMT) inattiva questi farmaci immunosoppressori aggiungendo un gruppo metilico. Negli individui con due alleli TMTP a bassa attività i farmaci sono metabolizzati lentamente; a dosaggi normali i farmaci si accumulano e possono causare una tossicità midollare gravissima
Emorragia	warfarin	*Vedi* sopra
Reazioni di ipersensibilità	abacavir, carbamazepina, allopurinolo	possono manifestarsi in modo marcato reazioni immunitarie inappropriate a farmaci altrimenti non tossici. Quando vengono trattati con il farmaco anti HIV abacavir, circa il 5% dei pazienti presenta reazioni da ipersensibilità cutanea, gastrointestinale e respiratoria che talvolta possono risultare fatali. Il trattamento con l'anticonvulsivante carbamazepina o con allopurinolo (usato per curare la gotta) può produrre reazioni farmacologiche avverse, compresa necrolisi epidermica tossica
Danno epatico	flucloxacillina, isoniazide, allopurinolo	i soggetti con determinati antigeni HLA sono maggiormente a rischio di malattia epatica indotta per alcuni farmaci, come l'*HLA-B*5701* nel caso della flucloxacillina.
Tossicità muscolare	alotano, isoflurano, varie statine	le statine e parecchi altri farmaci sono associati a miopatie solitamente di lieve entità. Talvolta, però, casi più gravi mostrano rabdomiolisi (distruzione del tessuto muscolare) che può avere esiti mortali. In risposta all'inalazione di anestetici (alotano, isoflurano), i soggetti che presentano mutazioni inattivanti nel gene del recettore della rianodina sviluppano una forma maligna di ipertermia caratterizzata da una rabdomiolisi gravissima e da un aumento estremo della temperatura (OMIM 145600)

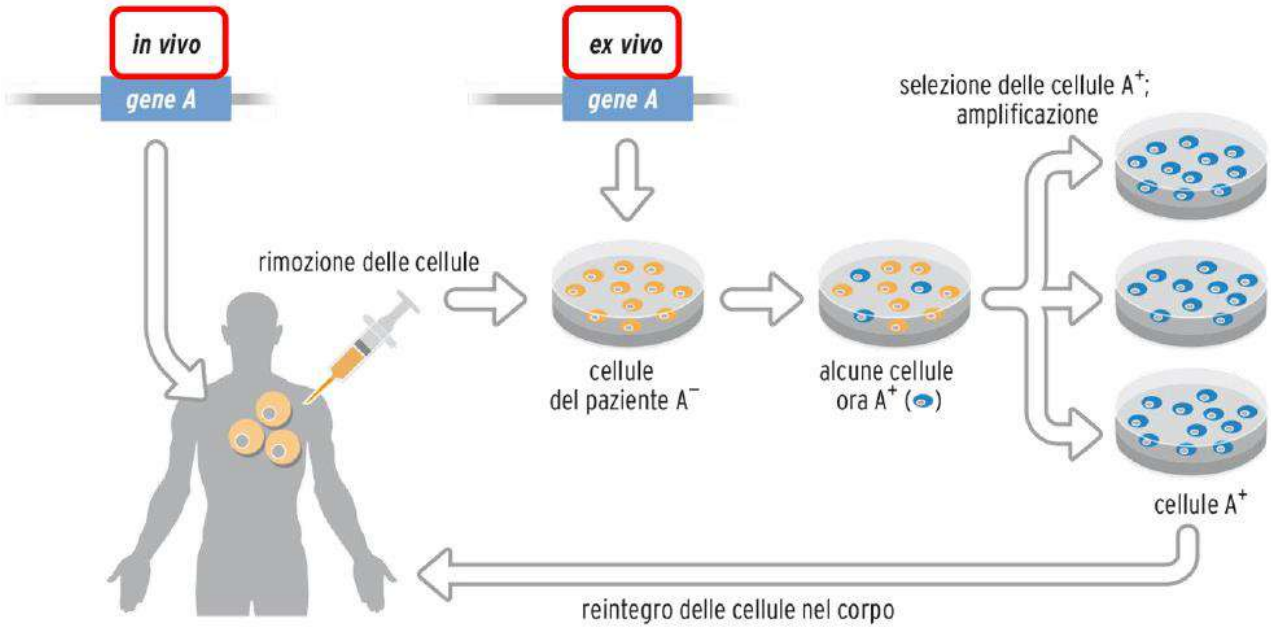
Ci sono differenti approcci genici nella terapia genica: l'addizione genica, l'uccisione diretta delle cellule attraverso l'inserzione di un gene suicida, il silenziamento genico, per esempio per bloccare la produzione di una proteina mutata, la riparazione del gene mutato, attraverso l'azione di un gene riparatore, l'uccisione indiretta delle cellule, metodo che prevede l'inserimento di un attivatore immunitario dentro cellule non patologiche, provocando una risposta immunitaria migliorata nell'organismo e, non appena le cellule hanno svolto il loro compito, viene indotta la loro morte.



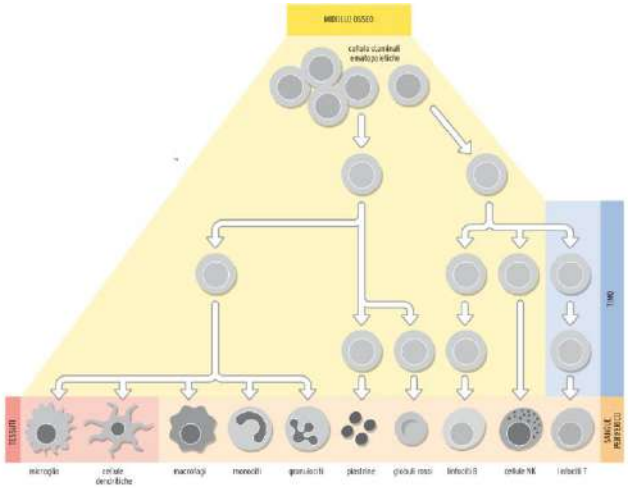
Di un trasferimento genico bisogna garantire efficienza e sicurezza. Esistono due sistemi di trasferimento del DNA in cellule umane, tramite virus e si chiama trasduzione (il virus, ad esempio un gammaretrovirus ricombinante terapeutico, che ha perso i propri geni patogeni, si può integrare oppure rimanere in forma episomale), non tramite virus e si chiama trasfezione.



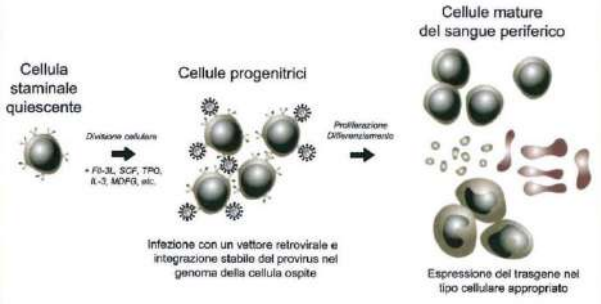
La terapia genica può avvenire in vivo, cioè iniettando direttamente il virus nel paziente, oppure ex vivo, cioè si estraggono cellule dal paziente, si inserisce il gene di interesse nelle cellule prelevate, si selezionano soltanto quelle trasformanti e si amplificano in coltura, dopodichè si reintegrano le cellule nel corpo del paziente.



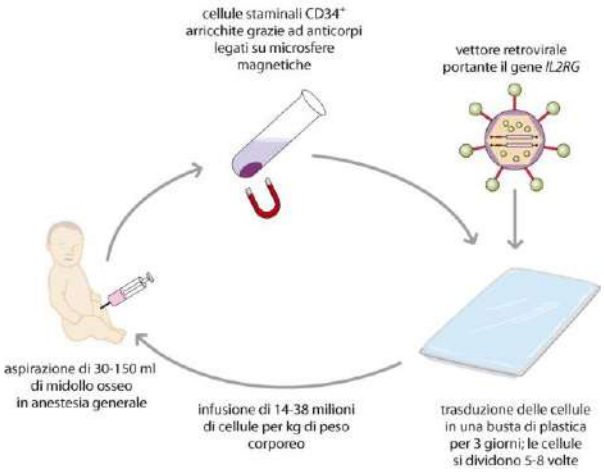
Un tipo di terapia ex vivo è stata applicata, con successo, a pazienti bambini affetti da X SCID. Dai piccoli vengono aspirati da 30 a 150 ml di midollo osseo in anestesia generale. Il midollo è ricco di cellule staminali ematopoietiche (HSC). Le cellule staminali CD34+ vengono arricchite grazie ad



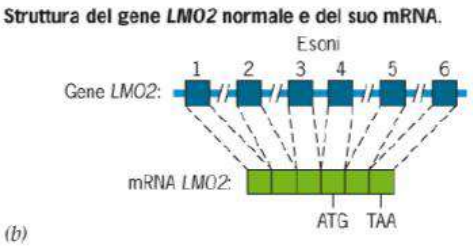
anticorpi legati su microsfere magnetiche. Si prepara un vettore retrovirale portante il gene gammaC. Le HSC sono cellule a riposo, che necessitano di essere stimolate con un cocktail di citochinine in grado di indurre il loro rientro nel ciclo cellulare.

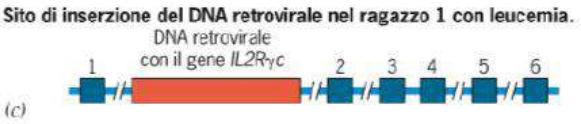


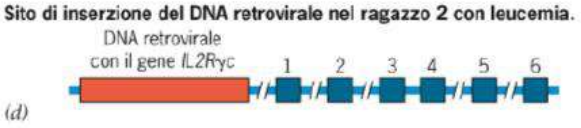
Dopo la stimolazione, le cellule sono trasdotte con il vettore retrovirale e lasciate trasformare per tre giorni, il tempo necessario per permettere la riproduzione delle cellule da 5 a 8 volte. Una volta moltiplicatesi, le cellule vengono reinfuse nel paziente in numero variabile da 14 a 38 milioni per kg di peso corporeo e qui si differenziano nei diversi stipiti cellulari che derivano da un'HSC, all'interno dei quali esprimono il gene con cui sono state trasdotte.



Si sono però verificati alcuni casi di oncogenesi provocata dall'inserzione di un retrovirus ingegnerizzato per combattere X-SCID. I pazienti svilupparono leucemie delle cellule T. I vettori si integrarono nei pressi di proto-oncogeni che vennero così attivati e portati ad esprimersi. Tre in particolare vennero identificati: LMO2, BMI1 e CCND2. Il gene LMO2 è composto da sei esoni. Nei due pazienti sottoposti alla terapia genica che hanno sviluppato la leucemia, il vettore retrovirale con IL2RGAMMAC si è integrato, in un paziente, tra gli esoni 1 e 2, mentre nell'altro a monte dell'esone 1. Il DNA retrovirale integrato ha aumentato i livelli di espressione del gene LMO2 nei globuli bianchi, nei quali normalmente l'espressione è repressa.







Con strategie analoghe si cura oggi anche l'ADA-SCID. La prima paziente sottoposta

a una terapia ex vivo nel 1990, all'età di 4 anni, Ashi DeSilva, da allora produce da sé le proprie cellule T funzionali.

Uno studio terminato nel 2009 si concentrò sull'efficacia e sulla sicurezza della terapia genetica, osservate a lungo termine, per la cura di ADA-SCID. Essa venne applicata su 10 bambini affetti per i quali mancava un donatore fratello o sorella con identico complesso maggiore d'istocompatibilità. Dopo 4 anni tutti i bambini erano ancora vivi. Le conclusioni tratte affermano che la terapia genica, combinata con un condizionamento a ridotta intensità, è un trattamento sicuro ed efficace per la SCID in pazienti affetti da assenza di ADA. Uno studio successivo, del 2016, stimò una sopravvivenza del 100% nei pazienti affetti da ADA-SCID trattati con cellule CD34+ geneticamente modificate. Furono osservati un attecchimento a lungo termine, la ricostituzione delle difese immunitarie e un minor numero di infezioni importanti in 15 dei 18 pazienti senza trasformazioni leucemiche.

La mancanza di ADA causa l'accumulo di metaboliti tossici che distruggono il sistema immunitario, provocando la SCID, spesso nota con il nome di malattia dei "bambini bolla". Strimvelis è una terapia genica approvata dall'Agenzia Europea per la Medicina pensata per pazienti affetti da ADA-SCID senza un donatore compatibile di midollo osseo, una situazione che si verifica in oltre il 75% dei casi. La malattia ha una frequenza di circa 1 su un milione, dunque di 10^-6. Fino al 2016 18 pazienti erano stati curati con successo.

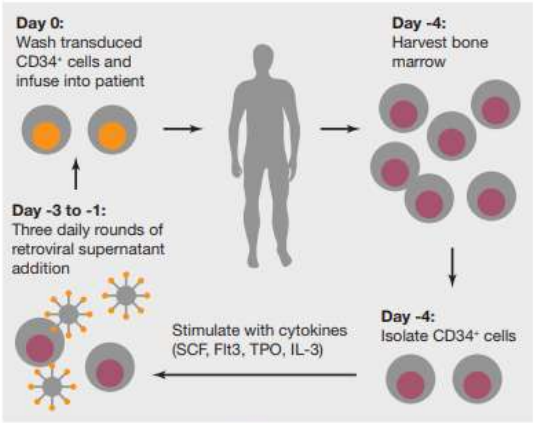
Leading Edge

Bench to Bedside

Cell

Gene Therapy for "Bubble Boy" Disease

Jonathan Hoggatt
Cancer Center and Center for Transplantation Sciences, Massachusetts General Hospital, Boston, MA 02129, USA
Correspondence: hoggatt.jonathan@mgh.harvard.edu
http://dx.doi.org/10.1016/j.cell.2016.06.049



Adenosine deaminase (ADA) deficiency results in the accumulation of toxic metabolites that destroy the immune system, causing severe combined immunodeficiency (ADA-SCID), often referred to as the "bubble boy" disease. Strimvelis is a European Medicines Agency approved gene therapy for ADA-SCID patients without a suitable bone marrow donor.

NAME
Strimvelis (GSK2696273); international nonproprietary name: autologous CD34+ enriched cell fraction that contains CD34+ cells transduced with retroviral vector that encodes for the human ADA cDNA sequence

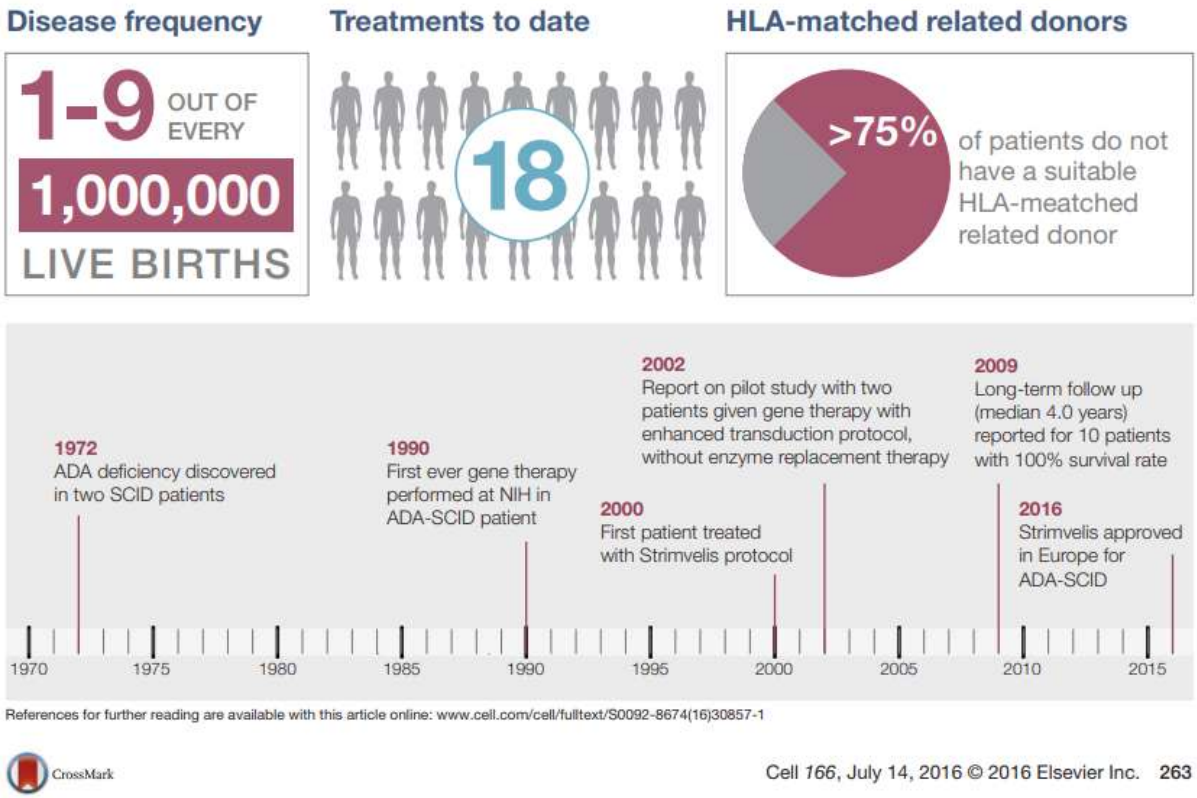
APPROVED FOR
Treatment of severe combined immunodeficiency due to adenosine deaminase deficiency (ADA-SCID) in patients who cannot be treated by a bone-marrow transplant because they do not have a suitable, matched, related donor

TYPE
Gene therapy: autologous CD34+ enriched cell fraction that contains CD34+ cells transduced with retroviral vector (LXSN vector) that encodes for the human ADA cDNA sequence

CELLULAR TARGETS
CD34+ hematopoietic cells

EFFECTS ON TARGETS
Replaces defective adenosine deaminase in immune cells

DEVELOPED BY
GlaxoSmithKline, MolMed, San Raffaele Telethon Institute for Gene Therapy

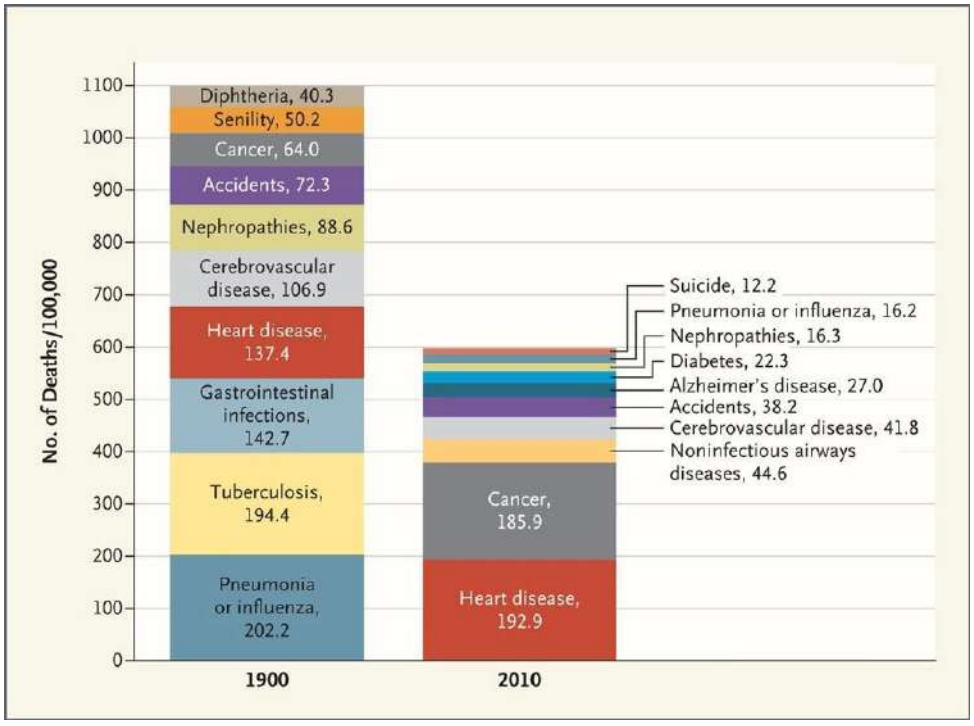


References for further reading are available with this article online: www.cell.com/cell/fulltext/S0092-8674(16)30857-1

CrossMark

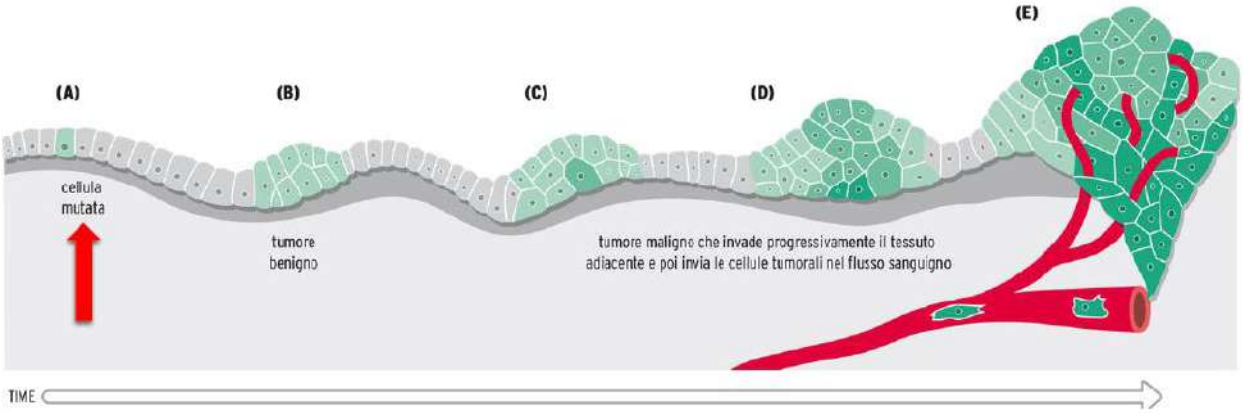
Cell *166*, July 14, 2016 © 2016 Elsevier Inc. 263

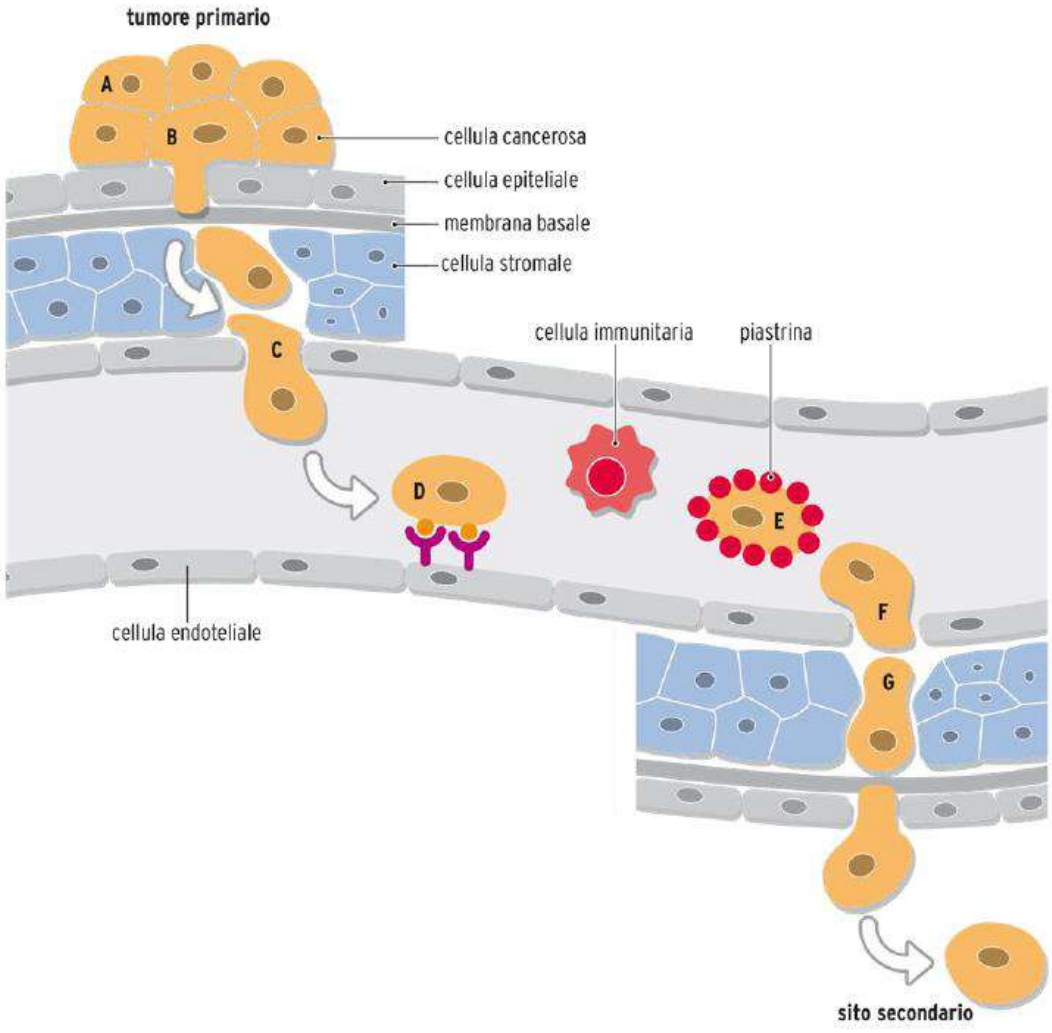
Genetica e genomica del cancro



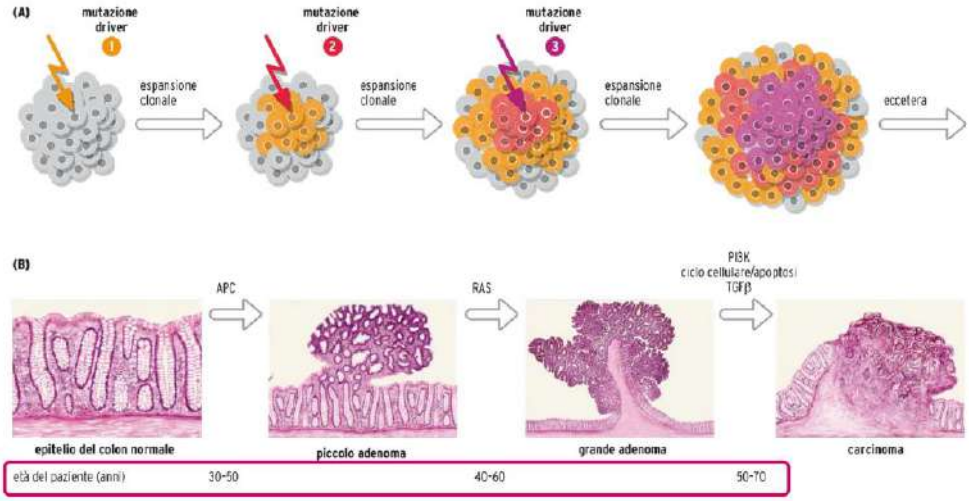
Nel 2010 il cancro venne stimato essere la seconda causa di morte nel mondo, subito dopo le malattie cardiovascolari. Il meccanismo con cui il tumore si sviluppa all'interno dell'organismo umano parte da una singola cellula mutata, che si replica un numero crescente di volte, formando un tumore benigno. Questa massa tumorale acquista un volume sempre maggiore fino ad essere considerata maligna. Invade progressivamente il tessuto adiacente e poi invia le cellule tumorali nel flusso sanguigno. Esse

colonizzano un altro tessuto e il tumore si estende ad una zona del corpo più ampia di quella iniziale. Le cellule cancerose si sviluppano a ridosso delle cellule epiteliali, attraversano la membrana basale e lo strato di cellule stromali e cellule endoteliali sottostante. Il sistema immunitario non riesce a riconoscere il pericolo perché le cellule cancerose vengono avvolte dalle piastrine che impediscono l'intercettazione. La cellula tumorale, dunque, può raggiungere indisturbato il sito secondario che colonizzerà.



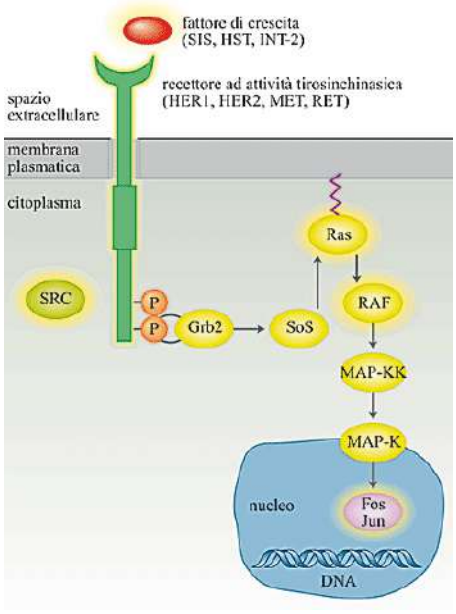


E' stata proposta un teoria "multistep" per la cancerogenesi: una cellula scelta casualmente viene colpita da una mutazione driver, cioè che porta la mutazione responsabile dell'insorgenza del tumore. La cellula inizia a replicarsi in maniera incontrollata, producendo diversi cloni di sé, anch'essi recanti la mutazione. La prima cellula mutata, già fuori controllo, può però accumulare altre mutazioni, essendo instabile, che renderanno ancora più aggressiva e rapida l'espansione clonale. Questo evento è destinato ad intensificarsi con l'avanzare dell'età, peggiorando le condizioni di salute del paziente e rendendo il tumore più difficile da rimuovere. Se consideriamo il caso di un epitelio del colon normale, ad esempio, una mutazione a carico di APC può provocare la nascita di un piccolo adenoma. Una seconda mutazione, a livello di RAS, ingrandirà l'adenoma di partenza. Infine, si verificheranno diverse mutazioni che coinvolgeranno fattori e proteine coinvolti nel ciclo cellulare e nei meccanismi di apoptosi e la forma assunta dalla massa tumorale sarà quella di un carcinoma.

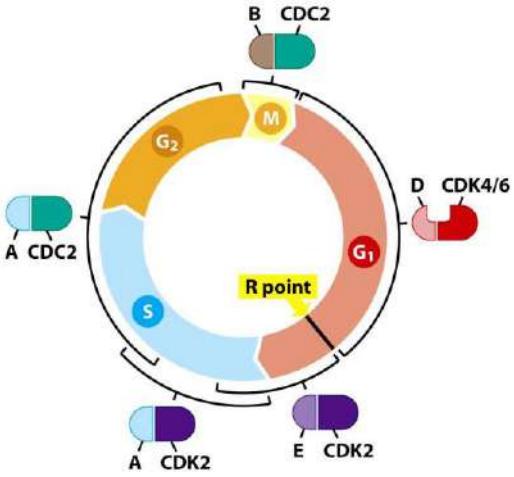


Nel DNA cellulare esistono i cosiddetti proto-oncogeni, geni inespressi normalmente che, se attivati, possono portare alla trasformazione di una cellula sana in cellula tumorale. Appartengono a diverse classi funzionali, che sono però tutte legate alla via di trasduzione del segnale per la proliferazione cellulare: sono fattori di crescita, recettori dei fattori stessi, fattori di trascrizione, regolatori del ciclo cellulare, inibitori dell'apoptosi.

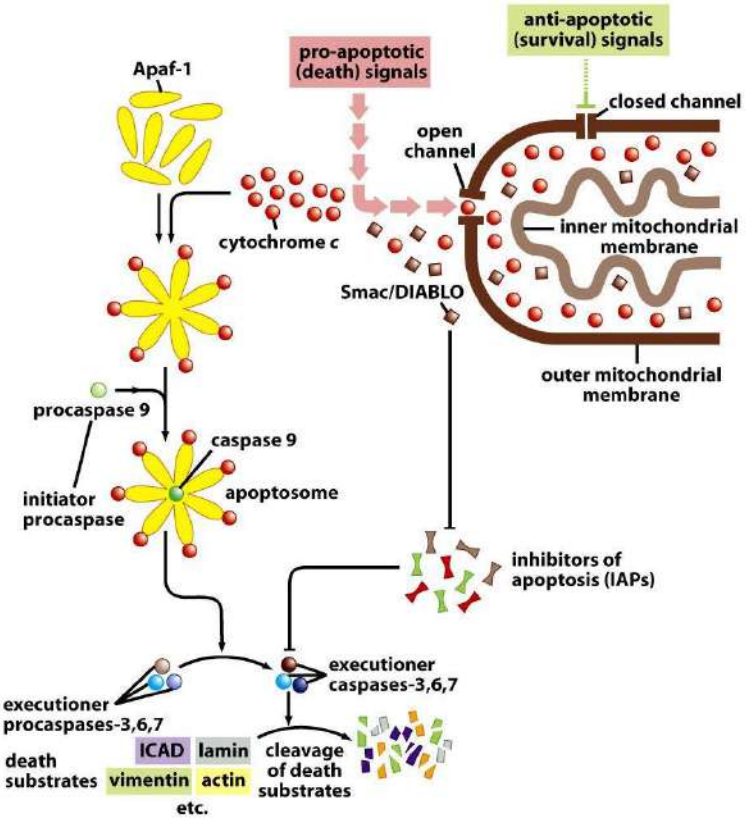
CLASSI FUNZIONALI	ESEMPI DI ONCOGENI	PRODOTTO DELL'ONCOGENE
Fattore di crescita secreto	*FGF4*	fattore di crescita dei fibroblasti di tipo 4
	PDGFB	fattore di crescita derivato dalle piastrine
Recettore del fattore di crescita; recettore della tirosin-chinasi	*EGFR*	recettore del fattore di crescita dell'epidermide
Componente delle vie di trasduzione del segnale	*ABL*	tirosin-chinasi citoplasmica
	HRAS	omologo dell'oncogene virale del sarcoma di ratto di Harvey, una GTPasi
	PIK3CA	subunità catalitica di una chinasi lipidica, fosfatidil inositolo chinasi
Fattore di trascrizione	*GLI1*	proteina *zinc-finger* che funziona come attivatore trascrizionale
	MYC	omologo dell'oncogene virale della mielocitomatosi. Attiva la trascrizione di vari geni correlati alla crescita
Regolatore del ciclo cellulare	*CCND1*	ciclina D
Inibitore dell'apoptosi	*BCL2*	proteina mitocondriale



Normalmente, quando una cellula riceve un fattore di crescita attraverso un recettore ad attività tirochinasica, si osserva una serie di reazioni a catena, che riguardano diverse proteine citoplasmatiche, che terminano nel nucleo, dove ha luogo l'attivazione della trascrizione dei geni bersaglio, specifici del ciclo cellulare. Proprio nel nucleo è contenuto un proto-oncogene c-Fos, che codifica una proteina, che forma un eterodimero con c-jun, coinvolto in una varietà di tumori, perché attivante un complesso AP-1 che si lega al DNA.

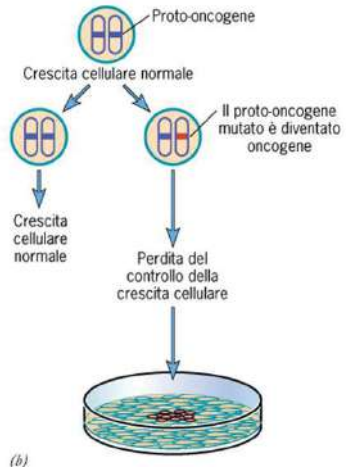


Un altro proto-oncogene importante, regolatore del ciclo cellulare, è CCND1, che ha come prodotto la ciclina D. Questa ciclina forma un complesso con CDK4 o CDK6 e in tale multiproteina costituisce una subunità regolatrice, la cui attività è funzionale alla transizione da G1 a S. E' stato dimostrato che interagisca con la proteina soppressore del tumore Rb e l'espressione di questo gene è regolata positivamente da Rb. Mutazioni, amplficazioni e overespressioni di questo gene, che alterino il naturale ciclo cellulare, sono frequentemente osservate in una varietà di tumori (es. tumore alla mammella) e contribuiscono alla cancerogenesi.

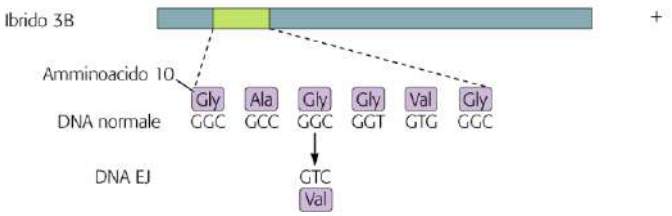


Un altro proto-oncogene responsabile di tumori è Bcl2, una proteina mitocondriale inibitrice dell'apoptosi. La cellula va incontro alla propria morte quando riceve un segnale pro-apoptotico extracellulare oppure si trova in una condizione di stress, trasmessa con segnali intracellulari. Questa seconda modalità, chiamata via intrinseca, si manifesta con disfunzioni mitocondriali, come la transizione di permeabilità mitocondriale, l'aumento di permeabilità della membrana esterna e il rilascio del citocromo c nel citosol. Il citocromo c si lega al complesso Apaf-1, che attiva le procaspasi 9, le forme inattive delle caspasi, e la loro unione produce un

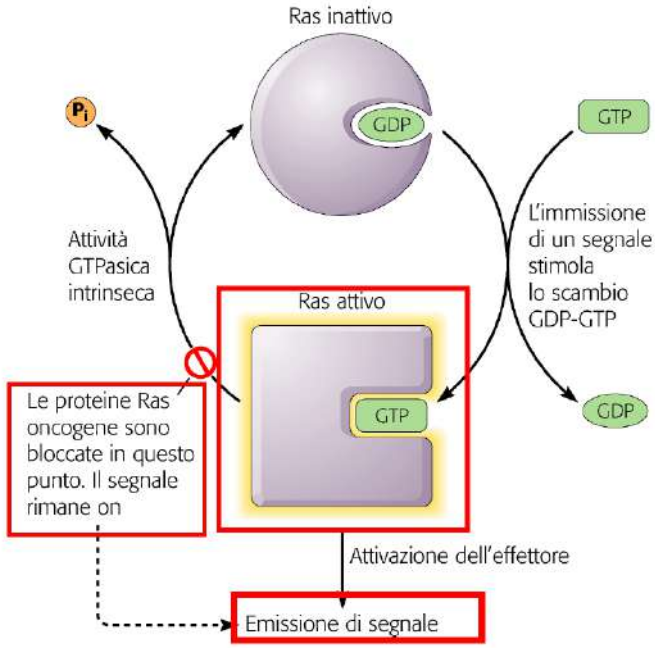
apoptosoma. Esso attiverà altre procaspasi dette "giustiziere", il cui target sono substrati cellulari che vengono sminuzzate e smaltite. Il compito di regolare la funzionalità dei mitocondri spetta alla famiglia bcl-2, che comprende, appunto, una serie di proteine (localizzate sulla membrana esterna) che regolano la permeabilità della membrana mitocondriale esterna e controllano il rilascio di citocromo c dai mitocondri. I membri di questa famiglia si dividono in due gruppi: pro-apoptotici e anti-apoptotici. Appartenente alla prima categoria è, ad esempio, il gene Bax, che stimola il rilascio del citocromo c nel citosol, garantendo il corretto svolgimento del processo di morte cellulare. Antagonista di Bax è il sopracitato Bcl2, che deve essere silenziato nel momento in cui la cellula riceve la sua "condanna a morte". Se però questo proto-oncogene si attiva, si verifica la chiusura dei canali che mettono in comunicazione lo spazio intermembrana, in seguito a un segnale di anti-apoptosi, cioè di sopravvivenza. Le cellule che esprimono notevolmente il gene bcl2 sono immortali e possono produrre alcuni tumori come ad esempio il linfoma o il cancro della prostata.



Per i proto-oncogeni si parla di gain of function, cioè basta che solo uno dei due alleli codificanti sia mutato perché si abbia l'attivazione della proteina, con conseguente perdita del controllo della crescita cellulare. Le modalità di attivazioni sono varie.

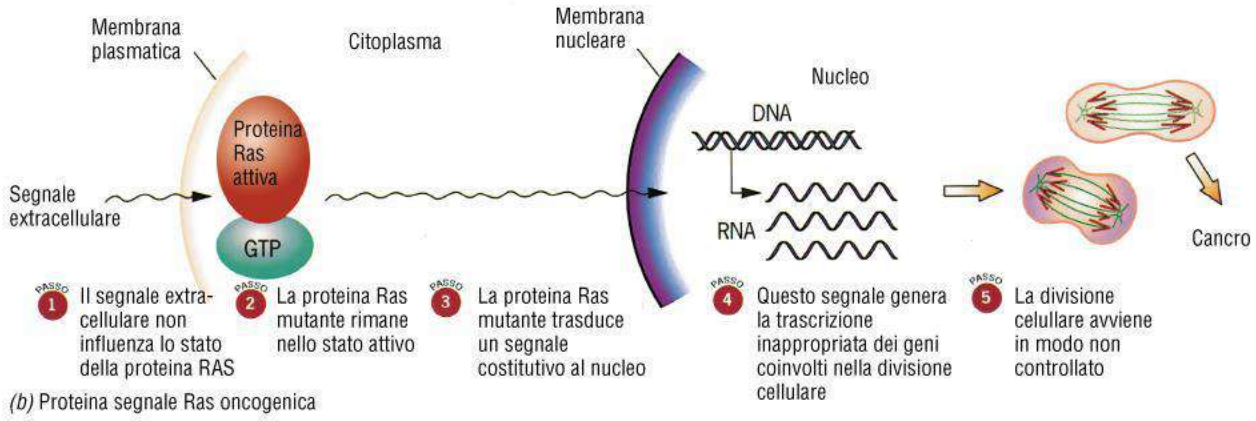


Possono avvenire mutazioni puntiformi, come nel caso del gene c-ras, dove una mutazione a carico dell'aa 12 (G>T) missenso, da Gly a Val, rende costitutiva la sintesi di questa proteina.



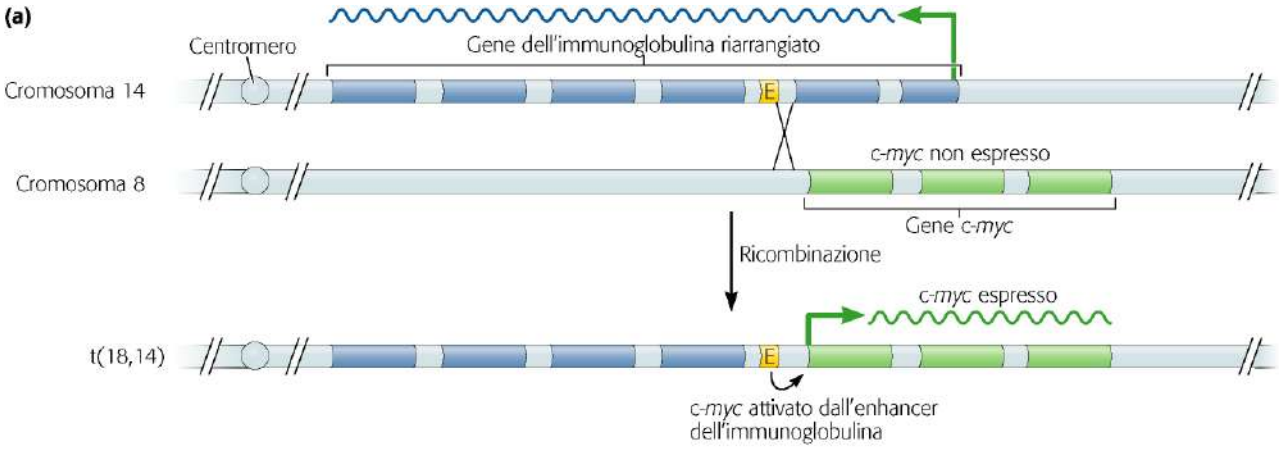
Essa è coinvolta nella via di trasduzione del segnale di proliferazione cellulare. Allo stato inattivo è legata ad una molecola di GDP. L'immissione di un segnale trasmesso stimola lo scambio GDP-GTP. Ras, allora, con un nuovo substrato GTP, è portata all'attivazione dell'effettore, con conseguente emissione di segnale. Una volta esaurito il proprio compito RAS, attraverso un'attività GTPasica intrinseca, torna nel proprio stato inattivo. Questa è la normale procedura. Le proteine Ras oncogene, invece, sono bloccate permanentemente

nello loro stato attivo, cioè viene loro impedito di esplicare l'attività GTPasica che appartiene loro. Si osserva, perciò, un'emissione di segnale costitutiva, cioè è sempre acceso l'impulso trasmesso al resto della catena. Da quel momento in poi la cellula rimane completamente indifferente ai segnali provenienti dall'esterno, che non riescono ad influenzare lo stato della proteina RAS mutante, che rimane dunque sempre nello stato attivo. Il segnale costituivo viene trasdotto al nucleo, dove genera la trascrizione inappropriata dei geni cellulari coinvolti nella divisione cellulare. La divisione cellulare avviene in modo non controllato, provocando l'insorgenza del cancro.

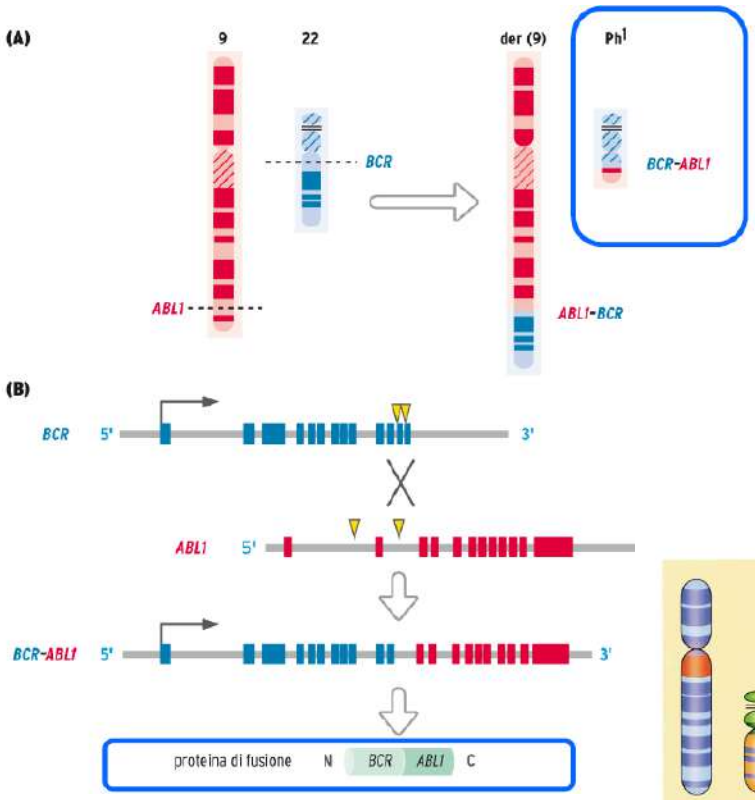


(b) Proteina segnale Ras oncogenica

Sono possibili anche traslocazioni cromosomiche, come quella che coinvolge i cromosomi 14 e 8. Dal cromosoma 14 vengono trascritti i geni coinvolti nella sintesi delle catene pesanti delle immunoglobuline, i linfociti B, dove avviene anche il fenomeno di riarrangiamento. Sul cromosoma 8 si trova invece il gene c-myc, normalmente non espresso. Quando avviene la traslocazione tra questi due cromosomi, l'enhancer dell'immunoglobulina attiva l'iperespressione del gene c-myc. Ciò avviene proprio nei linfociti B. Questa iperespressione induce le cellule a divenire cancerose.



Un'altra traslocazione dannosa per l'organismo avviene tra il cromosoma 9 (dove è situato il gene ABL1) e il 22 (dove è situato il gene BCR). La traslocazione reciproca fra questi due cromosomi provoca la formazione di un cromosoma associato alla leucemia mieloide cronica. I ricercatori ebbero la prova che la proteina di fusione è responsabile della leucemia: BCR-ABL veniva espressa nei topi



transgenici. I topi che erano stati colpiti da mutazioni missenso che distruggevano l'attività chinasica della proteina non avevano mai sviluppato leucemie.

Era chiaro che il target terapeutico era l'attività chinasica di ABL. Le chinasi hanno strutture diversificate nello stato inattivo e bisognava pensare ad un farmaco la cui conformazione calzasse la forma attiva di BCR-ABL1. Venne sviluppato il Gleevec o imatinib mesilato. Questo farmaco si inserisce perfettamente nel sito attivo della chinasi e blocca il loop di attivazione in una conformazione inattiva. I risultati furono eclatanti: con le terapie convenzionali il 50% dei pazienti sopravviveva 5 anni, mentre l'89% dei pazienti trattati con Gleevec vive 5 anni e più (una piccola percentuale di tumori sviluppa una resistenza al Gleevec).

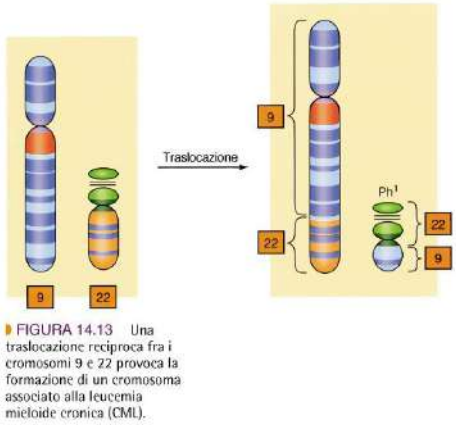
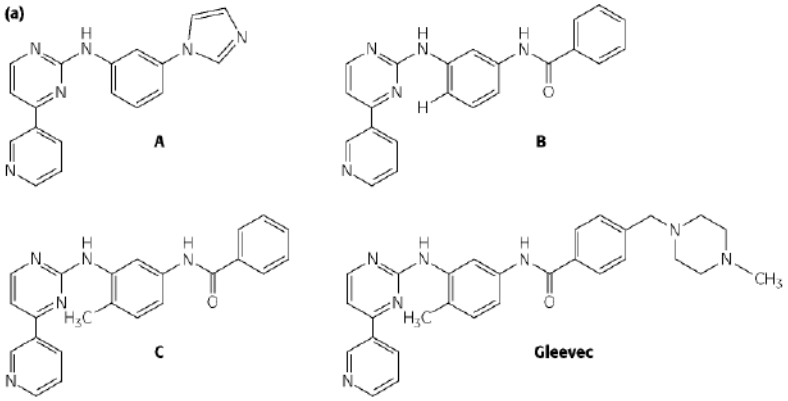
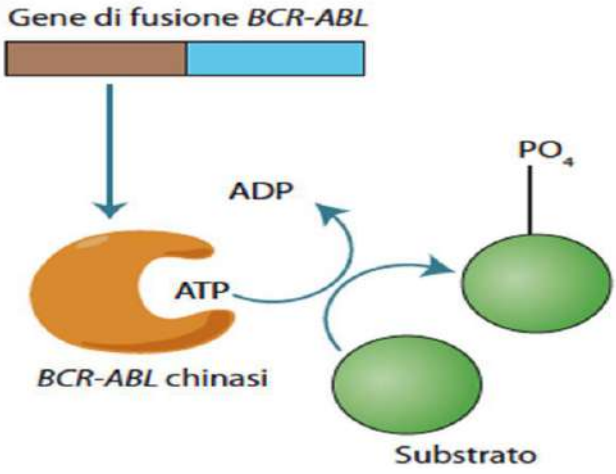
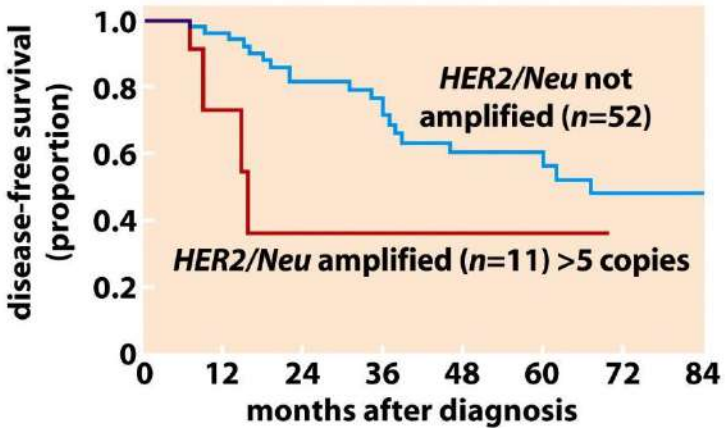


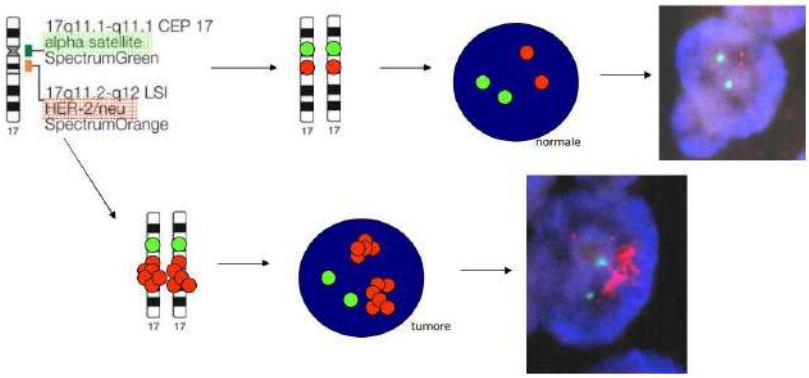
FIGURA 14.13 Una traslocazione reciproca fra i cromosomi 9 e 22 provoca la formazione di un cromosoma associato alla leucemia mieloide cronica (CML).

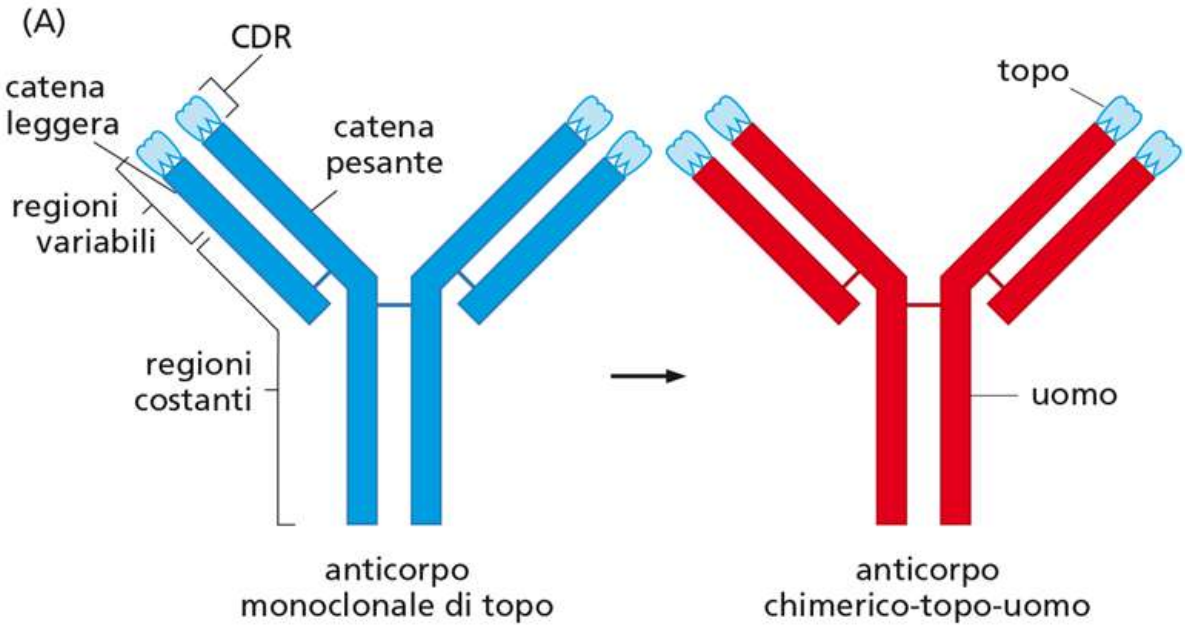




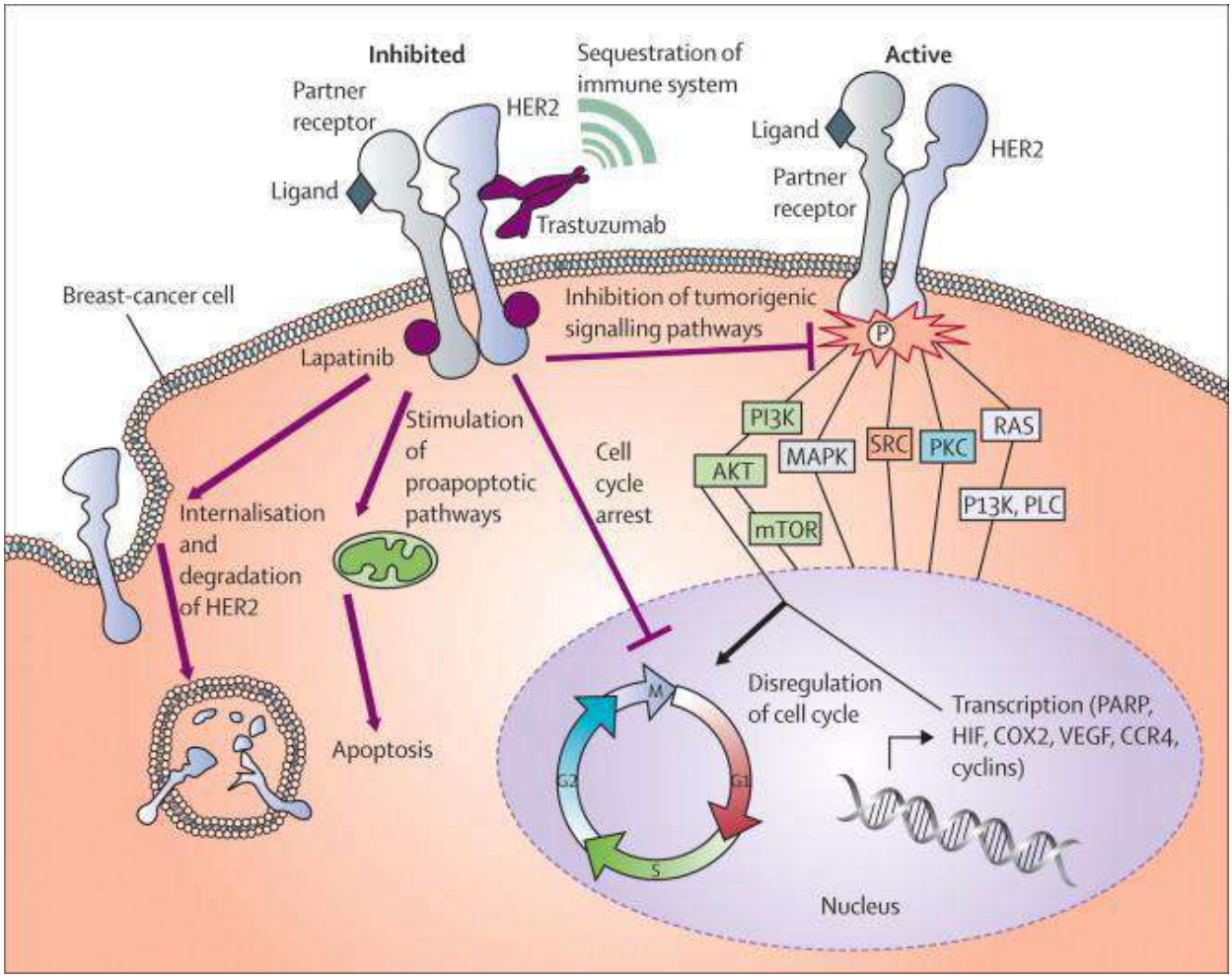
Il terzo caso di attivazione di proto-oncogeni è costituito dall'amplificazione genica, un fenomeno per cui, durante la replicazione del DNA, non si ottiene una sola copia di una regione cromosomica, come sarebbe corretto, ma molte di più. La cellula si ritrova così ad avere più copie di uno stesso gene, presenti come piccoli cromosomi indipendenti (*double minutes*) oppure come inserzioni nei cromosomi normali (*HSRs* - regioni a colorazione omogenea). In molti tipi di cancro l'amplificazione genica riguarda gli oncogeni. Questa disfunzione riguarda, ad esempio, il gene HER2. Il gene HER2 codifica per una proteina superficiale di membrana c-erbB-2 che forma degli eterodimeri con altri membri della famiglia (HER1, HER3 e HER4) in risposta a ligandi specifici detti neuregoline/heregoline. Tale dimerizzazione determina l'attivazione di una cascata di segnali che comportano un aumento della proliferazione e differenziazione cellulare. Tale attivazione si traduce in una iper-espressione del gene HER2 a livello dei nuclei interfasici facilmente identificabile mediante la tecnica di ibridazione in situ fluorescente. L'iper-espressione di HER2, comunemente chiamata amplificazione genica, si osserva nel 20-30% dei tumori mammari e in minor percentulae nei tumori gastroesofagei e dello stomaco.







Per combattere questa forma di tumore è stata favorita la produzione di anticorpi geneticamente modificati, in questo caso di un anticorpo chimerico-topo-uomo, dove la maggior parte della struttura è di origine umana, ma i siti di legame per l'antigene provengono dal topo e legano il dominio extracellulare di erbB2 con alta affinità. Questo sistema è stato commercializzato sotto il nome di erceptina (o Trastuzumab), un farmaco molto potente nella lotta contro il tumore alla mammella, capace di ridurre dell'80% il numero di cellule staminali del cancro. Si parla di immunizzazione passiva, perché la cellula si difende dal rischio di sviluppare tumori inattivando il proprio sistema immunitario.



Una categoria altrettanto importante e antitetica agli oncogeni è quella degli oncosoppressori. Una differenza tra questi due gruppi è costituita dal diverso effetto che provocano le mutazioni in loro. Mentre una mutazione in uno dei due alleli di un oncogene può essere sufficiente per indurre la cellula a perdere il controllo sulla crescita, entrambe le copie di un gene onco-soppressore devono essere inattivate per avere lo stesso effetto. In effetti, gli oncogeni derivano dai proto-oncogeni come risultato di mutazioni con guadagno di funzione, vale a dire mutazioni che inducono il prodotto genico ad esibire delle nuove funzioni che portano allo stato maligno. Al contrario, i geni onco-soppressori subiscono mutazioni con conseguente perdita di funzione e/o inattivazione epigenetica che li rendono incapaci di frenare la crescita cellulare.

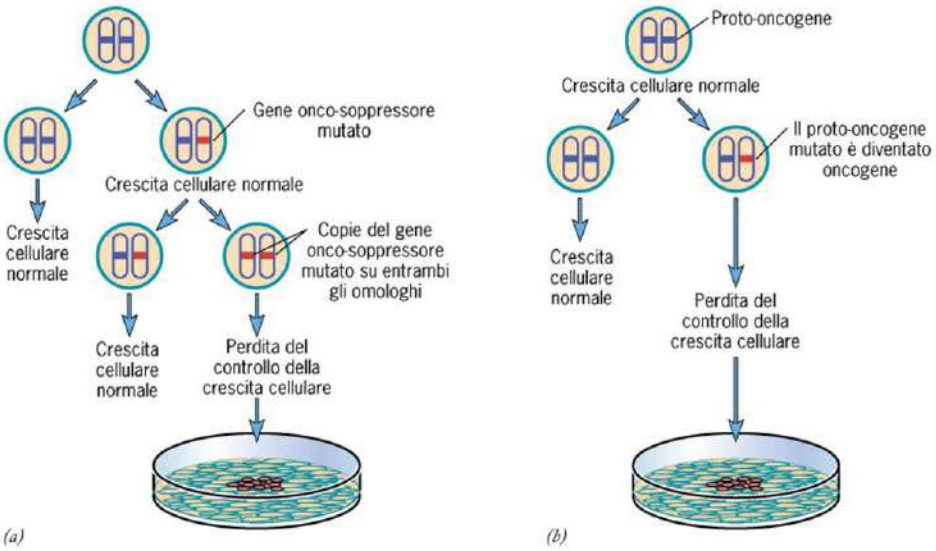


FIGURA 16.8 Effetti contrastanti di mutazioni in geni onco-soppressori (*a*) e in oncogeni (*b*). Mentre una mutazione in una delle due copie (alleli) di un oncogene può essere sufficiente per indurre la cellula a perdere il controllo sulla crescita, entrambe le copie di un gene onco-soppressore devono essere inattivate per avere lo stesso effetto. Come si vedrà in seguito, gli oncogeni derivano dai proto-oncogeni come risultato di mutazioni con guadagno di funzione, vale a dire mutazioni che inducono il prodotto genico ad esibire delle nuove funzioni che portano allo stato maligno. Al contrario, i geni onco-soppressori subiscono delle mutazioni con perdita di funzione e/o inattivazione epigenetica che li rendono incapaci di frenare la crescita cellulare.

Un noto oncosoppressore è RB, il cui substrato è il fattore E2F e il cui malfunzionamento provoca il retinoblastoma. Quando sono legati, significa che non è necessario trascrivere i geni indispensabili per l'entrata in fase S. In seguito alla ricezione di un segnale di un fattore di crescita, si attiva il complesso Cdk-ciclina G1, che interagisce con un ATP (che diventa ADP). Questa proteina fosforila Rb, che cambia conformazione e si distacca E2F, permettendo la trascrizione e la traduzione di enzimi e altre proteine necessari per l'entrata in fase S.

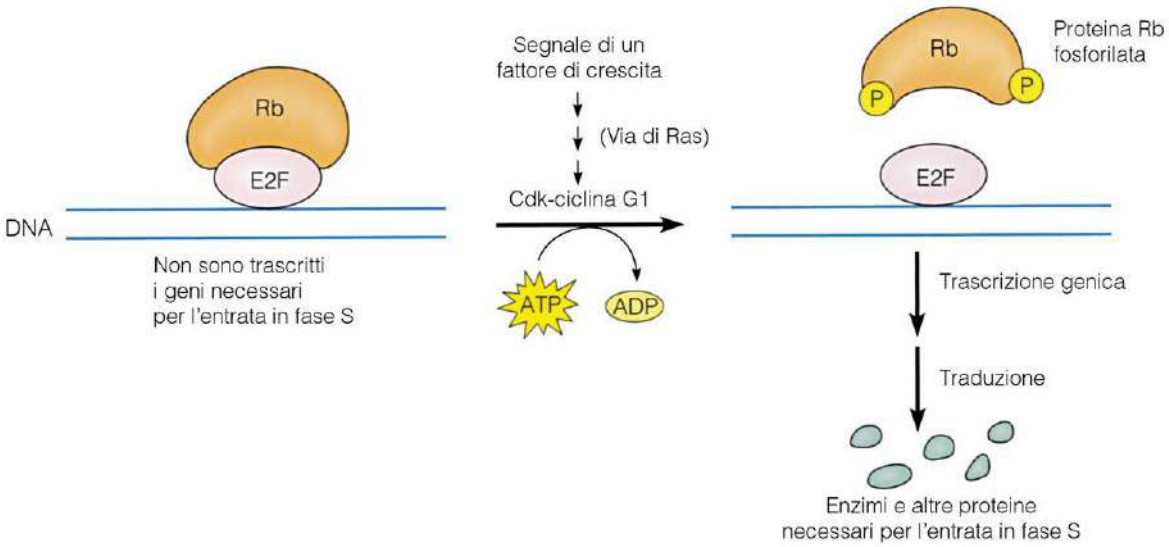
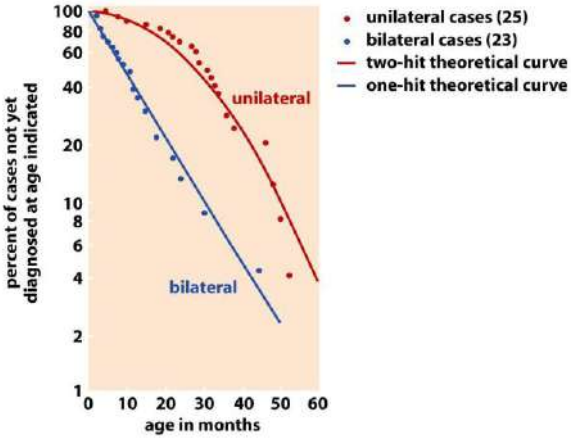


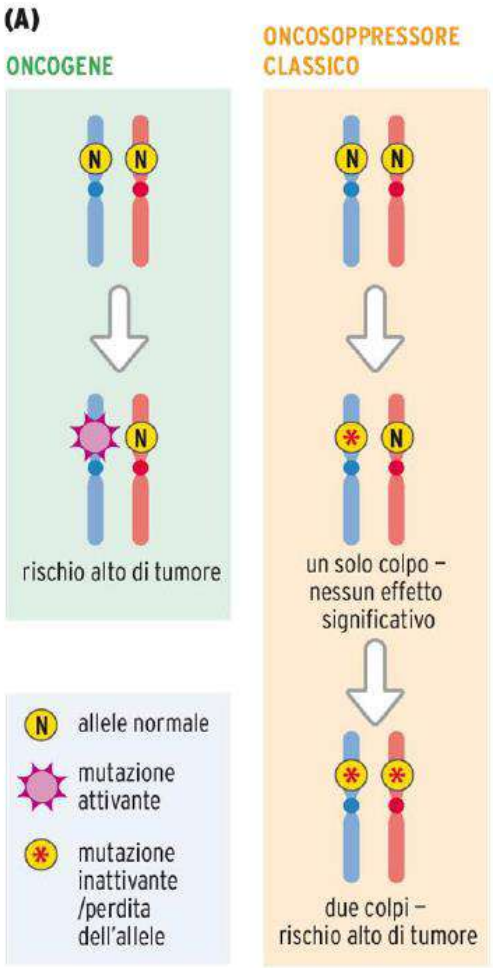
Figura 17-36

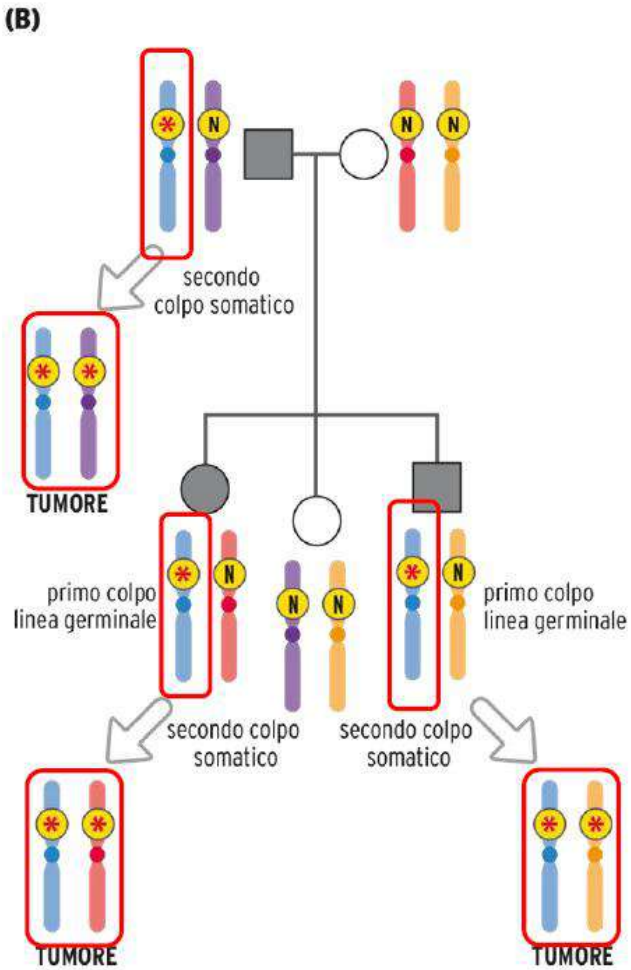
In base ai dati raccolti negli anni '70 è stata formulata l'ipotesi di Knudson (1971) o "ipotesi dei due colpi". Era noto che il retinoblastoma, una malattia rara, può presentarsi in forma familiare e in forma sporadica. La prima mostra caratteristiche compatibili con una trasmissione autosomica



dominante e ha una penetranza variabile da famiglia a famiglia. Inoltre il RB può essere uni- o bilaterale. Quando è unilaterale, possono essere presenti diverse aree separate di cellule neoplasiche, indicative della presenza di tumori multipli indipendenti. Knudson notò che le forme familiari e le forme bilaterali mostravano caratteristiche diverse rispettivamente da quelle sporadiche e quelle unilaterali.

Le forme bilaterali, come quelle familiari, tendono a insorgere in età più precoce. Inoltre, i pazienti affetti da forme familiari hanno molto spesso tumori bilaterali, mentre solo una frazione limitata dei casi sporadici presenta tumori in entrambi gli occhi oppure più aree tumorali in un singolo occhio. Knudson partì dal presupposto che la presenza di forme ereditarie è indicativa dell'intervento di varianti costituzionali che predispongono allo sviluppo del tumore, e ipotizzò che lo stesso tipo di alterazioni fosse coinvolto nell'origine dei tumori sporadici. Il difetto di penetranza (variabile) indica però che la mutazione costituzionale non è sufficiente a far comparire la malattia, dimostrando quindi la necessità di altri eventi per lo sviluppo del RB nei soggetti predisposti. Per spiegare le differenze osservate, egli formulò la teoria dei "due colpi" (two hits), secondo la quale lo sviluppo del RB richiede due eventi mutazionali.

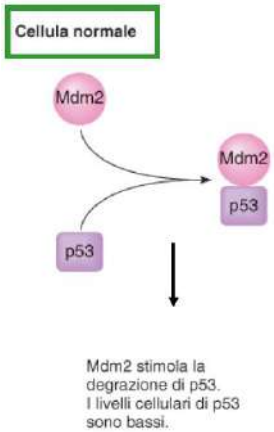




Nelle forme familiari una prima mutazione è presente in tutte le potenziali cellule bersaglio (i precursori dell'epitelio differenziato della retina), poiché l'individuo l'ha ereditata per via germinale da uno dei due genitori, eterozigote per la stessa mutazione, oppure perché si è verificata una mutazione germinale de novo; le conseguenze sono uguali nei due casi (mutazione ereditata o de novo), poiché la mutazione, essendo presente già allo stadio di zigote, farà parte del genotipo costituzionale dell'individuo. Quando in una cellula somatica si verifica il secondo evento mutazionale, inizia lo sviluppo del tumore.

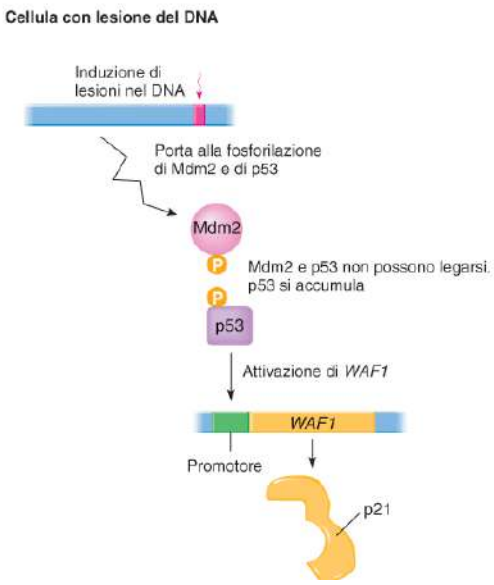
Nelle forme non ereditarie entrambe le mutazioni devono avvenire a livello somatico. Perché si sviluppi il RB, esse devono essere contemporaneamente presenti nella stessa cellula; le cellule in cui si è verificata una sola mutazione, così come le loro discendenti, hanno caratteristiche normali; acquisiranno proprietà tumorali solo se e quando comparirà nel loro genoma la seconda mutazione. Questa ipotesi spiega perché il rischio di RB è in genere molto più basso negli individui con storia familiare negativa (forme sporadiche) rispetto a coloro che hanno ereditato un gene mutante, per i quali la probabilità di ammalarsi equivale alla penetranza, quindi 0,9. Nei soggetti presisposti basta una sola mutazione in una qualsiasi cellula bersaglio, mentre i tumori sporadici compaiono solo quando si sono verificate indipendentemente due mutazioni, la seconda delle quali deve avvenire in una cellula in cui è già presente la prima.

Un altro oncosoppressore, la cui proteina è implicata nel mantenimento dell'integrità genomica, è TP53. Davanti al riconoscimento di danni subiti a carico al DNA p53 provoca l'arresto del ciclo cellulare oppure induce l'apoptosi nella cellula colpita.

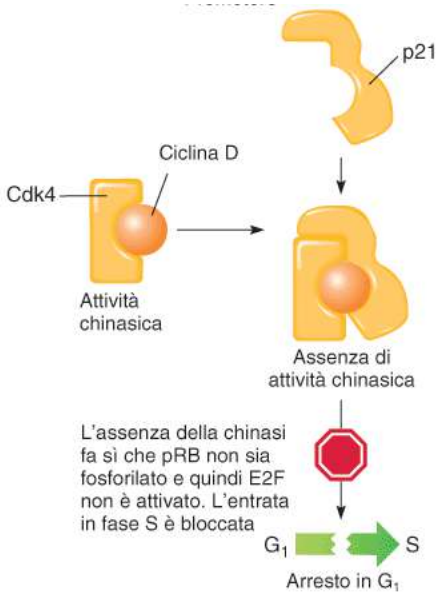


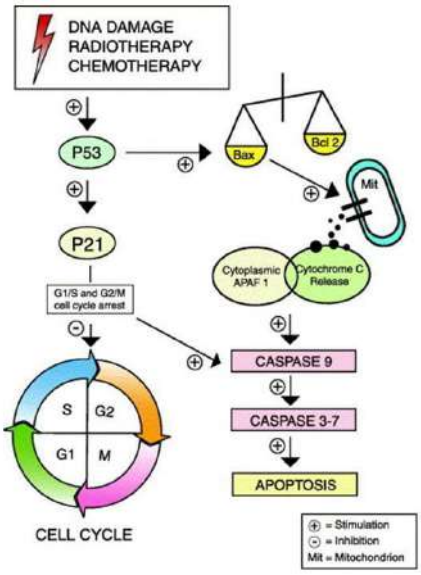
Il suo gene mappa sul cromosoma 17 braccio corto ed è formato da 11 esoni. Si compone di due domini di attivazione trascrizionale, un dominio di proteine, un dominio di legame al DNA, un sito di localizzazione nucleare, un dominio di oligomerizzazione.

In una cellula normale, p53 si lega a Mdm2, il suo principale inibitore, che ne stimola la degradazione, inducendone l'ubiquitinazione e abbassandone dunque i livelli cellulare. Se invece si inducono lesioni del DNA, si osserva la fosforilazione di Mdm2 e di p53, che, in questa nuova conformazione non possono legarsi tra di loro: p53 si accumula. La proteina, presente in gran quantità nella cellula, allora si lega al

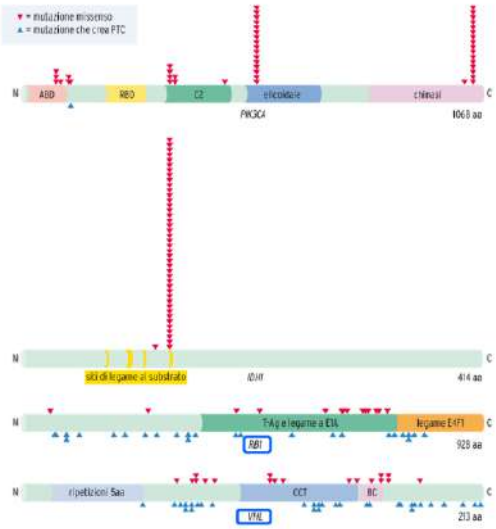


promotore del gene WAF1, attivandone la trascrizione. La proteina che ne deriva, p21, si lega al complesso Cdk4+Ciclina D, attivo tra la fase G1 e la fase S, inibendone l'attività chinasica e provocando l'arresto del ciclo in G1, in attesa che i danni vengano riparati.





Se i danni registrati sono particolarmente gravi ed estesi, p53 induce l'apoptosi a livello dei geni pro-apoptotici, favorendo la sintesi della proteina Bax a discapito di Bcl2, che quindi induce l'apertura dei canali mitocondriali e il rilascio del citocromo c nel citosol.



Altri geni bersaglio di mutazioni missenso oppure di mutazioni che producono codoni di terminazione prematuri sono PIK3CA, IDHI, RB1, VHL. La localizzazione di questi eventi è leggibile nei cosiddetti spettri di mutazione.

Le mutazioni a carico del gene di p53 sono dominanti negative. Ciò significa che una sola copia mutata annulla la funzionalità del gene, perché una sola della subunità (4 totali) è malformata e impedisce il corretto assemblamento del complesso proteico. Se le subunità

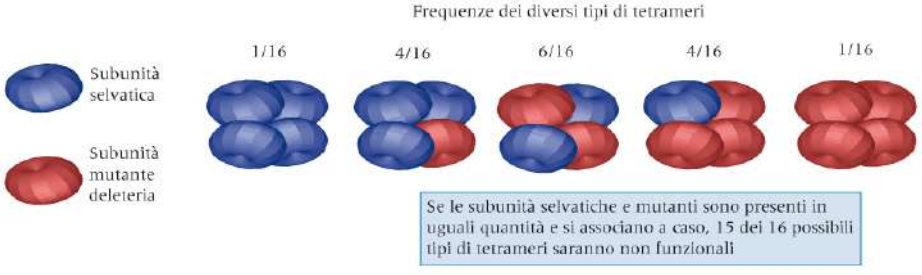
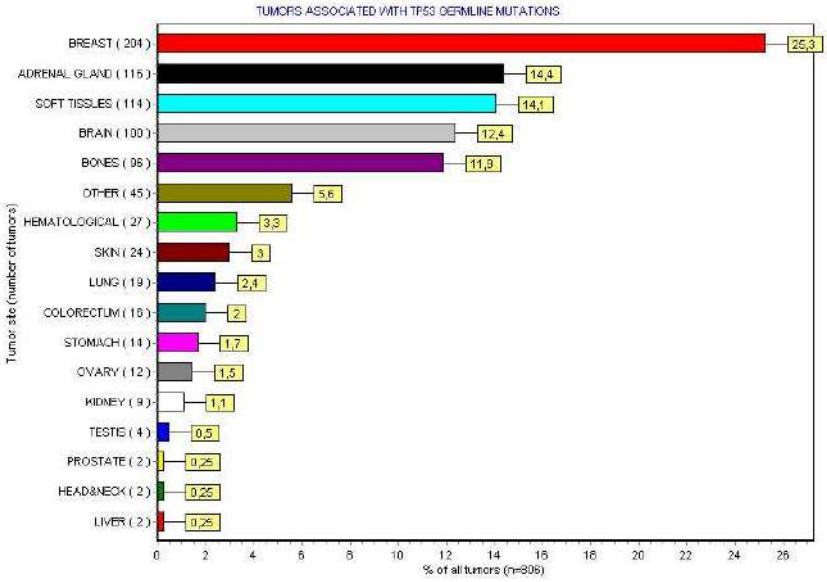
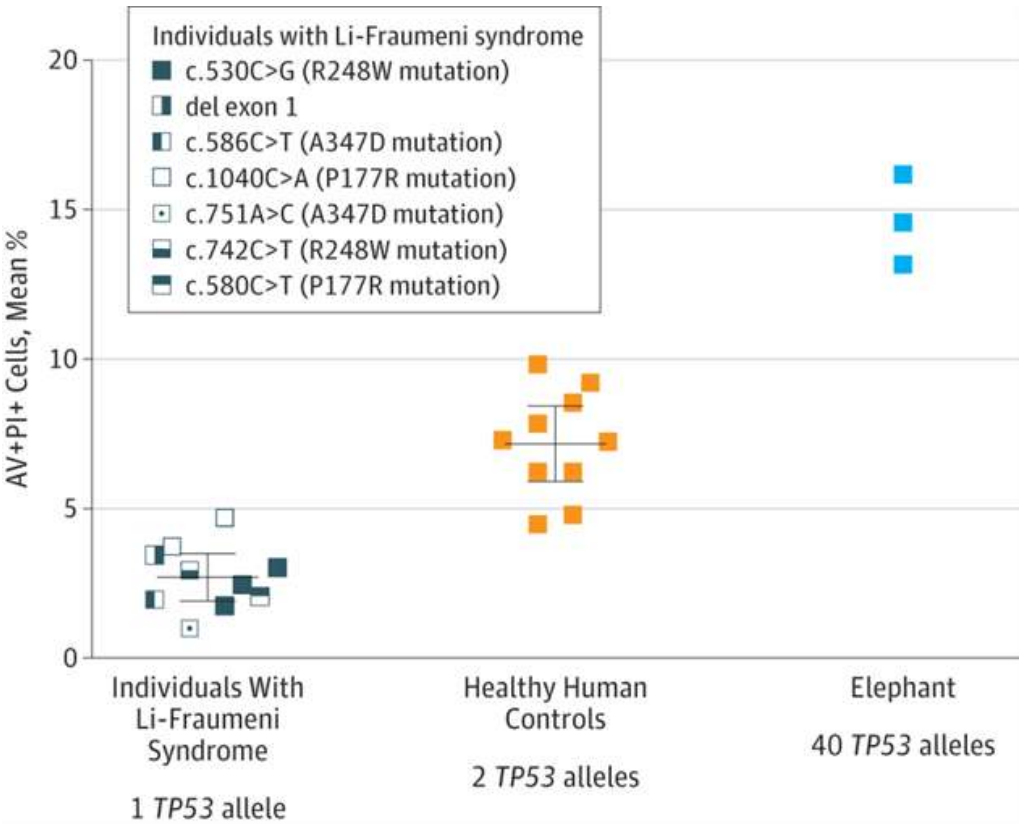


FIGURA 15.28 Le conseguenze sulla funzione di p53 di una mutazione dominante-negativa del suo gene.



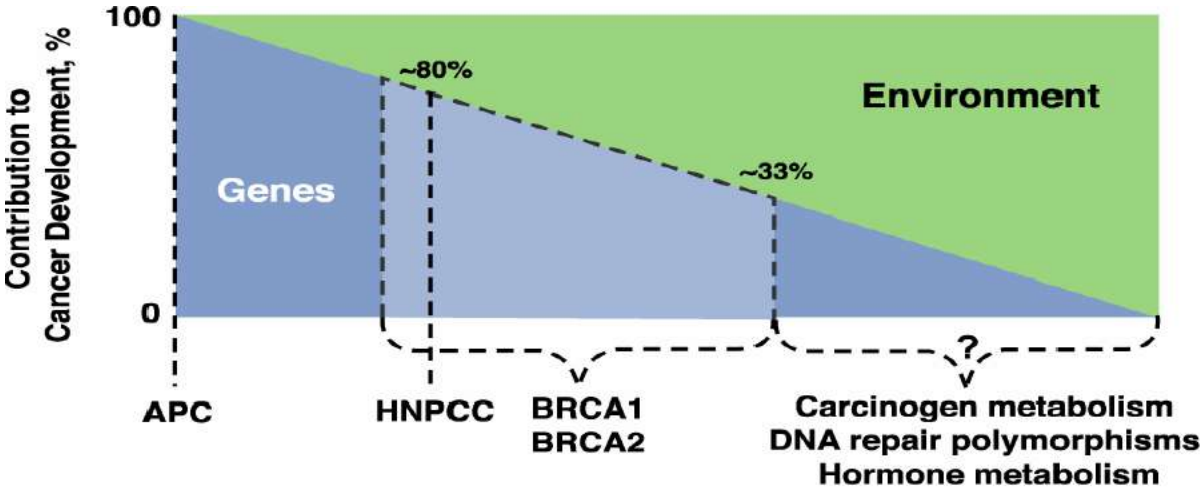
selvatiche e mutanti sono presenti in uguale quantità, 15 dei possibili tipi di tetrameri saranno non funzionali.

Mutazioni a carico di TP53 nella linea germinale sono responsabili di tantissimi tipi di tumori o malattie che rendono il paziente più predisposto a sviluppare forme tumorali, come la sindrome di Li-Fraumeni (nella quale le mutazioni avvengono nel dominio di legame al DNA, che dunque non viene correttamente riconosciuto).



Uno studio del 2015 studiò i potenziali meccanismi responsabili della resistenza al cancro negli elefanti e confrontò la risposta cellulare in questi animali rispetto all'essere umano davanti a un danno al DNA. Queste furono le conclusioni: confrontati con altre specie di Mammiferi, sembra che gli elefanti abbiano un tasso di tumori più basso di quello atteso, probabilmente collegato a un diverso numero di copie di T'53 (40 alleli). Rispetto alle cellule umane, le cellule degli elefanti hanno dimostrato una risposta di tipo apoptotica più frequente in seguito a un danno al DNA indotto (il doppio rispetto a un individuo umano sano e tre volte tanto rispetto a un paziente affetto da Li-Fraumeni, dopo esposizione a radiazioni ionizzanti). Queste scoperte, se replicate, potrebbero

rappresentare un approccio, basato sull'evoluzione, per comprendere i meccanismi legati alla soppressione del cancro.



Ma qual è l'effetto dell'ambiente sull'insorgenza di tumori? In quale misura contribuisce allo sviluppo di queste malattie? In alcuni casi sappiamo che non gioca alcun ruolo, in altri influisce fino a un 70%, ma rimangono irrisolte le interazioni tra ambiente e, ad esempio, metabolismo degli oncogeni o polimorfismi di riparazione del DNA o metabolismo degli ormoni. Gli effetti dell'ambiente registrati sono stati individuati tramite studi sui gemelli e su diverse popolazioni (basandosi sugli emigrati).

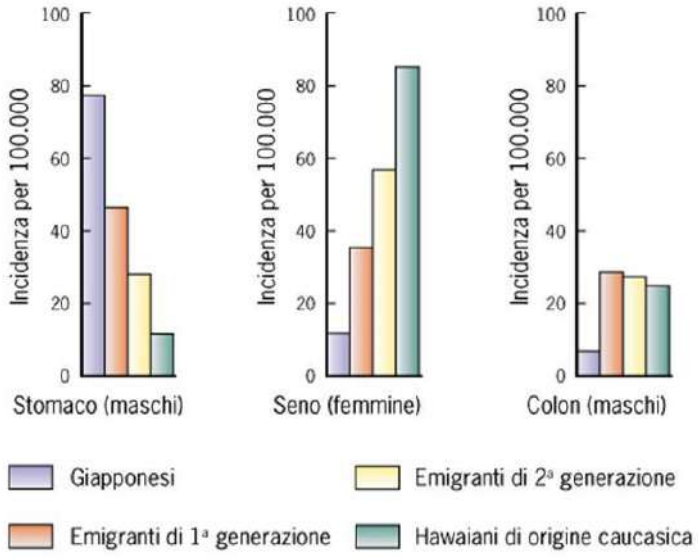


FIGURA 16.6 Variazione dell'incidenza di cancro in individui di origine giapponese in seguito alla migrazione nelle Hawaii. L'incidenza di cancro dello stomaco diminuisce, mentre quella di cancro della mammella e del colon aumenta. Comunque, dei tre tipi di cancro, solo quello del colon ha raggiunto tassi equivalenti a quelli degli Hawaiani caucasici alla seconda generazione. (DA L. N. KOLONEL, ET AL. PER GENT. CONC. DA NATURE REVS. CANCER 4:3, 2004; © COPYRIGHT 2004, MACMILLAN MAGAZINES LTD).

Uno studio del 2004 indagò l'incidenza di cancro in individui di origine giapponese in seguito alla migrazione nelle Hawaii. L'incidenza di cancro allo stomaco diminuiva rispetto ai connazionali rimasti in patria e si avvicinava al tasso dei nativi hawaiani, mentre quella di cancro alla mammella e al colon aumentava. Comunque, dei tre tipi di cancro, solo quello del colon per gli emigrati della seconda generazione ha raggiunto tassi equivalenti a quelli degli Hawaiani caucasici (autoctoni).

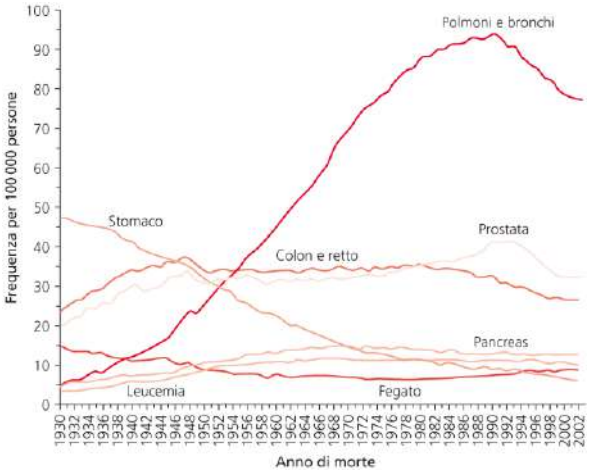
Diversi sono i fattori di rischio, che influiscono negativamente sulla salute degli individui, favorendo l'insorgenza di tumori.

Il primo è l'esposizione occupazionale, che venne studiata per la prima volta a fine '700 (Potts 1775) tra gli spazzacamini, che, a continuo contatto con la fuliggine, sviluppavano tumori allo scroto. In tempi più recenti, gli operai addetti all'asfaltatura delle strade sono spesso esposti a idrocarburi

Fig. 2.19 Asphalt road-workers (shown here in India) are exposed to polycyclic aromatic hydrocarbons.

Fig. 2.21 Textile dyeing in Ahmedabad, India. Protection against occupational exposures is often suboptimal in developing countries.

aromatici ciclici, pesantemente nocivi. Anche le operazioni di tintura tessile devono essere svolte in totale sicurezza a causa dei solventi usati. Altre sostanze dannose, presenti in ambienti di lavoro, sono il benzene e il cloruro di vinile (CVM), che può provocare un angiosarcoma al fegato. L'amianto causò numerosi casi di tumori alle vie respiratorie e non solo fino al 1992, anno in cui fu emanata una legge che ne proibiva l'uso.



Il tabacco, principalmente inalato con il fumo di sigarette, è un altro fattore di rischio, in particolare per i tumori ai polmoni e ai bronchi. Il fumo di sigaretta, infatti, è ritenuto responsabile dell'80% dei nuovi casi di tumore, a causa del grande numero di sostanze aggiunte al filtro che sono deleterie per l'organismo umano, come: butadiene, benzene, stirene, formaldeide, acetaldeide, IPA, N-Nitrosammina, ammine aromatiche, metalli pesanti. La maggior incidenza del tumore ai polmoni si registra in Europa, soprattutto nella parte orientale, in Nord America e in Australia.

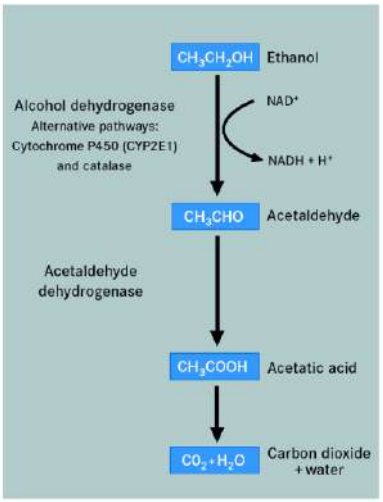


Fig. 2.12 The major pathway of alcohol metabolism in humans.

L'alcol, terzo fattore di rischio, provoca, ad esempio, numerosi casi di tumori orali tra la popolazione maschile. I tassi più alti appartengono a India, Australia, Francia, Ungheria, Brasile e Sudafrica. Il consumo di alcol, inoltre, aumenta il rischio di cancro del tratto gastrointestinale superiore. Il problema fisiologico si attesta al livello del metabolismo dell'alcol dove avviene la formazione di acetaldeide, che è, in quantità ingenti, dannosa per il corpo umano. Particolarmente deleteria per l'uomo si conferma la sinergia alcol-fumo, connessa allo sviluppo di tumori alla laringe.

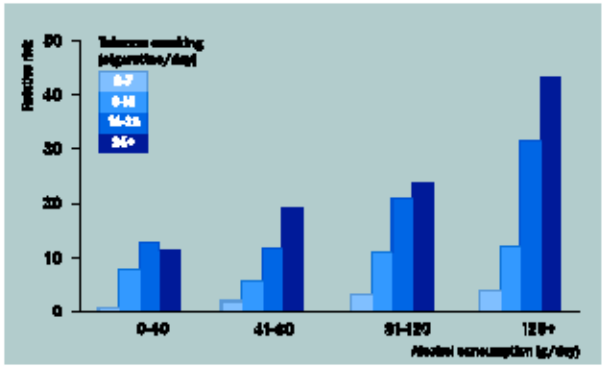


Fig. 2.14 Multiplicative increase in relative risk of laryngeal cancer as a consequence of both alcohol drinking and active smoking (colour coding approximates progressive doubling of risk as exposure increases). A.J. Tuyns et al. (1988) *Int J Cancer*, 41:483-91.

Studi dimostrano una connessione tra l'insorgenza di tumori ai polmoni e l'esposizione a inquinanti atmosferici, principalmente provenienti da industrie e da emissioni di automobili. Mentre in alcune parti dell'Asia si registra un alto tasso di incidenza di cancro ai polmoni tra le donne a causa del carburante usato per scaldare e cucinare e alti livelli di vapori tossici di oli da cottura.

Fig. 2.24 Automobile emissions are a major source of atmospheric pollution.

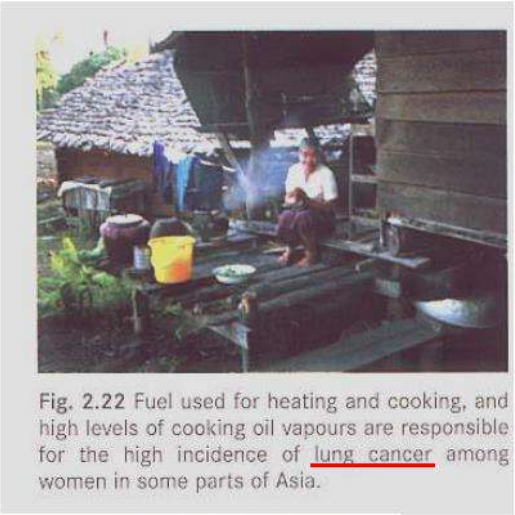


Fig. 2.22 Fuel used for heating and cooking, and high levels of cooking oil vapours are responsible for the high incidence of lung cancer among women in some parts of Asia.

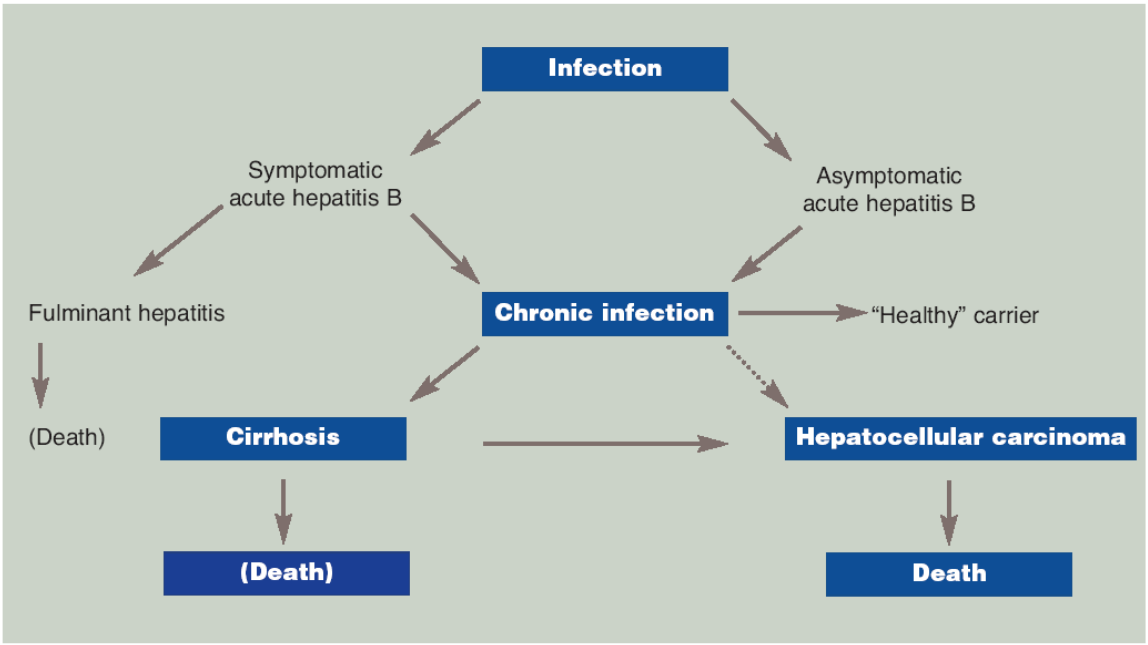


Fig. 8.04 Overview of outcome of HBV infection.

Alcune infezioni croniche, come il virus dell'epatite B e C (HBV e HCV), indeboliscono l'organismo e "preparano il terreno" per lo sviluppo di forme tumorali. In seguito ad un'infezione si possono sviluppare due tipi di epatite B: quella sintomatica e quella asintomatica. La prima può originare un'epatite fulminante, potenzialmente mortale, oppure rimanere cronica nell'organismo, come fa la seconda. Il paziente diventa così portatore "sano" di epatite B, se non si verificano ulteriori complicazioni. L'infezione latente, infatti, può svilupparsi in una cirrosi epatica, o, più raramente, in

un carcinoma epatico. La cirrosi può condurre alla morte oppure evolversi in un carcinoma, che molto spesso ha esito letale. Il microrganismo Helicobacter pylori, abitante del nostro stomaco, invece è stato collegato alla formazione di tumori allo stomaco, scoperta che è valsa il Nobel per la Medicina nel 2005 per Marshall. Il tumore del collo dell'utero è strettamente connesso al papillomavirus umano (HPV), scoperta di zur Hausen che gli valse il Nobel per la Medicina nel 2008. La sua maggiore incidenza si registra in America Centrale, Africa del Sud, India.

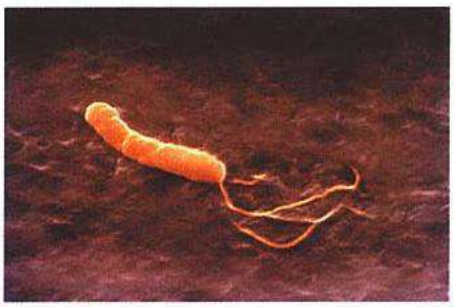


Fig. 2.45 The *Helicobacter pylori* bacterium structure as revealed by scanning electron microscopy.

L'80% di questi Paesi è in via di sviluppo. Correlazioni tra gli HPV e il tumore al collo dell'utero vennero sospettate fin dagli anni '70.

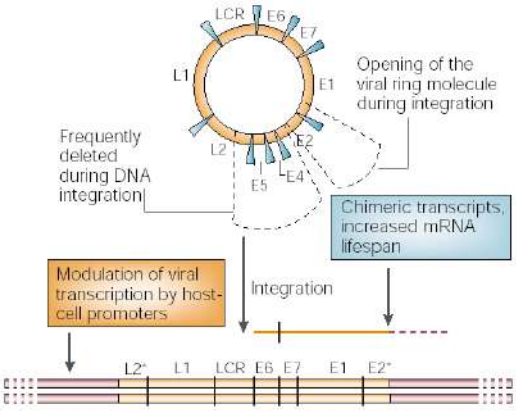


Figure 2 | **The organization of circular HPV DNA and its integration into host-cell DNA.** The human papillomavirus (HPV) genome contains between 6800 and 8000 base pairs and is divided into eight open reading frames — E6, E7, E1, E2, E4, E5, and L2 and L1 — coding for 'early' (E) or 'late' (L) functions. In the course of cancer development, the viral molecule frequently becomes integrated into host-cell DNA. The ring molecule is most often opened within the E2 open reading frame, disrupting the continuity of that gene. Part of E2 and open reading frames that are adjacent to E2 — E4, E5 and part of L2 — are regularly deleted after integration (partial genes are represented by an asterisk). Viral transcripts, which uniformly span the E6 and E7 region, and are often linked to flanking cellular sequences, are present and transcription might be modulated (enhanced) by flanking host-cell promoters. LCR, long control region.

DNA di tipi specifici di HPV sono stati trovati da allora in quasi tutte le biopsie per il tumore alla cervice. Gli oncogeni dell'HPV che sono espressi in quelle cellule sono coinvolti nella loro trasformazione e immortalità e sono richiesti per la progressione verso la malignità. Studi epidemiologici hanno sottolineato che gli HPV sono il principale fattore eziologico per lo sviluppo del tumore alla cervice. Il DNA circolare del virus si apre e si integra (con 8 reading frame) in una cellula umana della cervice, una parte dell'anello viene deleta e la trascrizione del DNA viene modulata dai promotori della cellula ospite. Si ottengono trascritti chimerici, che formano mRNA con un arco di vita più ampio. Osservazioni sperimentali hanno dimostrato che E6interagisce con *P53*, e E7con *RB* bloccando l'attivita di questi "tumour suppressor genes". Il tipo di HPV prevalente in questo tipo di tumori è il 16; altri diffusi sono il 6 e l'11.

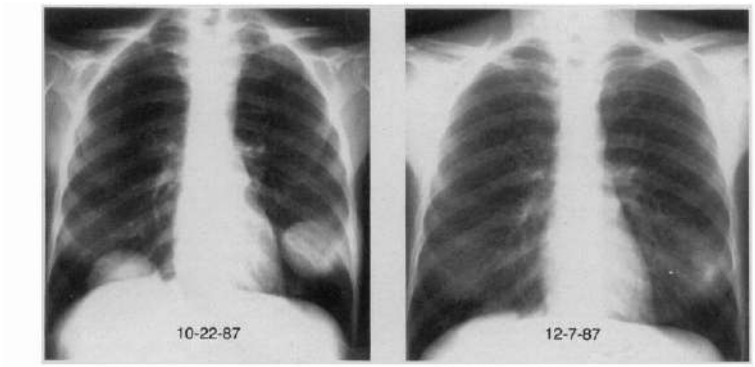


Figure 1. Pulmonary Metastases in Patient 4 before (Left) and after (Right) One Course of TIL, Interleukin-2, and Cyclophosphamide. The later chest film shows regression of the metastases. Multiple subcutaneous metastases regressed completely as well.

Già dagli anni'80 si pose il problema di come trattare i tumori con la terapia genica. Il primo lavoro fu di Steven Rosenberg (1988), che decise di trattare 20 pazienti affetti da melanoma allo stadio metastatico con TIL (Tumor-Iinflitrating Lymphocytes) e interleuchina 2. Egli

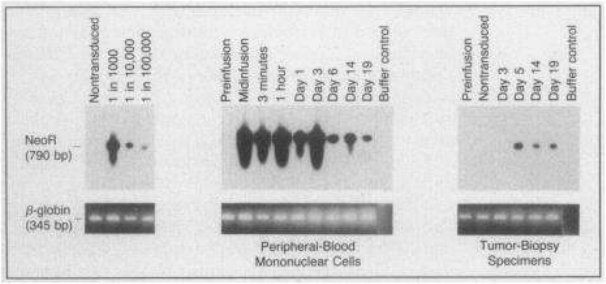
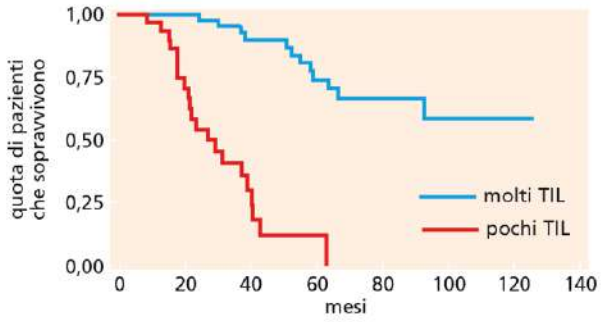


Figure 3. Polymerase-Chain-Reaction Detection of the NeoR Gene in Peripheral-Blood Lymphocytes and Lymphocytes Grown from Tumor-Biopsy Specimens from a Patient Infused with Gene-Transduced TIL.

Samples of genomic DNA were amplified by the polymerase chain reaction with synthetic oligonucleotide primers for the NeoR gene and the human beta-globin gene. The amplified beta-globin gene was detected by homidium bromide staining, and the NeoR gene by Southern blot analysis. To semiquantitate the results of the polymerase-chain-reaction assay, serial 10-fold dilutions of NeoR-transduced, G418-selected TIL were mixed with nontransduced TIL, and the assay was run in parallel with the samples from the patient.

osservò che nei pazienti sottoposti a questa terapia, nuova per l'epoca e ancora in fase sperimentale, si registrava una risposta più efficace rispetto ai risultati ottenuti con la somministrazione della sola interleuchina 2 o con cellule killer attivate con linfochina.

Era però troppo presto per determinare se quella nuova forma di immunoterapia potesse migliorare la sopravvivenza di chi era affetto dal melanoma, ma sicuramente ulteriori prove sembravano giustificate.

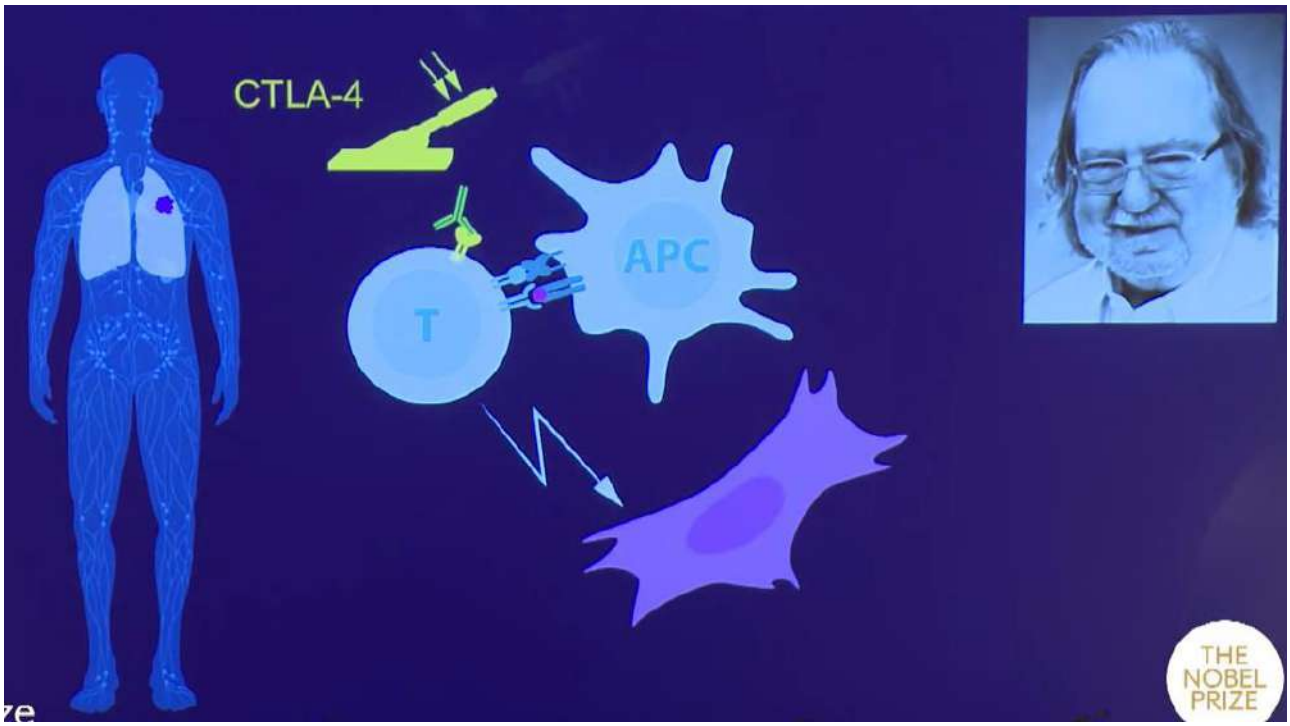


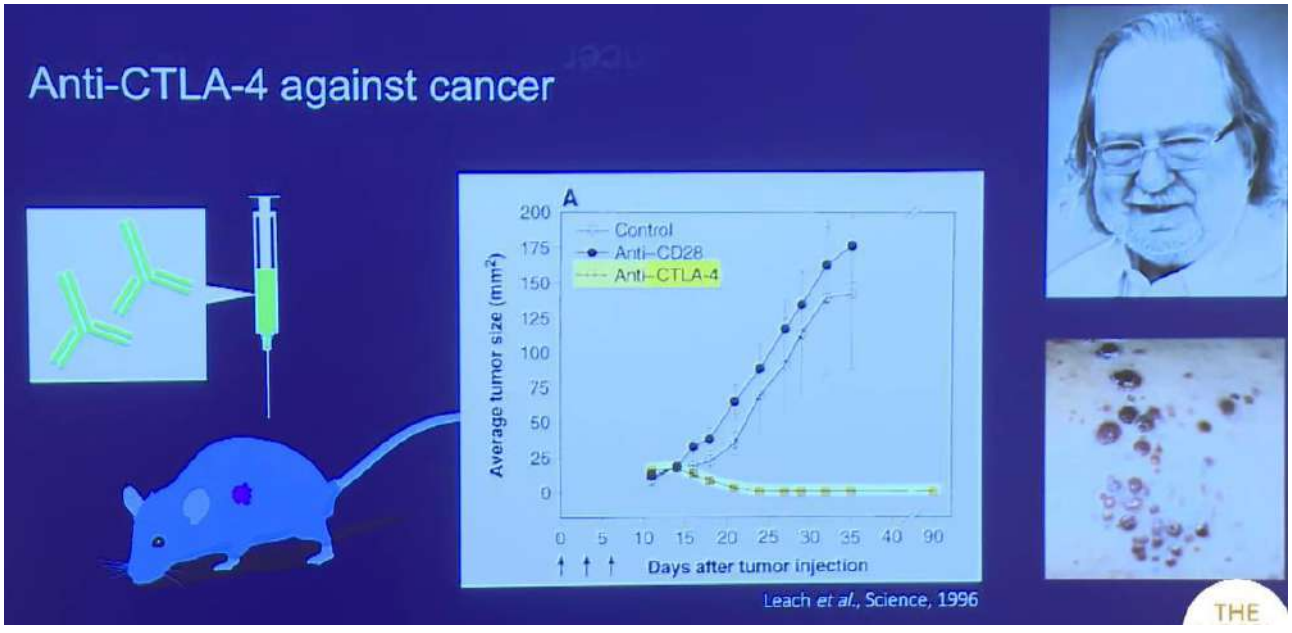
Due anni più tardi Rosenberg effettuò un'operazione analoga, ma aggiunse un elemento in più. Si era osservato che il trattamento con TIL e interleuchina 2 provocava la regressione del melanoma metastatico in circa metà dei pazienti. Per ottimizzare la terapia e affinare la distribuzione e la sopravvivenza dei TIL, optarono per una trasduzione genica mediata da un retrovirus per introdurre il gene per la resistenza alla neomicina nei TIL umani prima dell'infusione nei pazienti. Questi studi dimostrarono la possibilità e la sicurezza dell'uso di una trasduzione genica retrovirale per la terapia genica umana e ebbero implicazioni sulla progettazione di TIL con potenza antitumorale incrementata, come pure sul possibile uso dei linfociti per la terapia genica di altre malattie. Fu il primo esperimento, che diede poi il via a gene-tracking e immunoterapia, dove un essere umano riceveva materiale genetico esogeno. L'anno successivo sempre Rosenberg proseguì nell'applicazione dell'immunoterapia e della terapia genica contro il cancro.

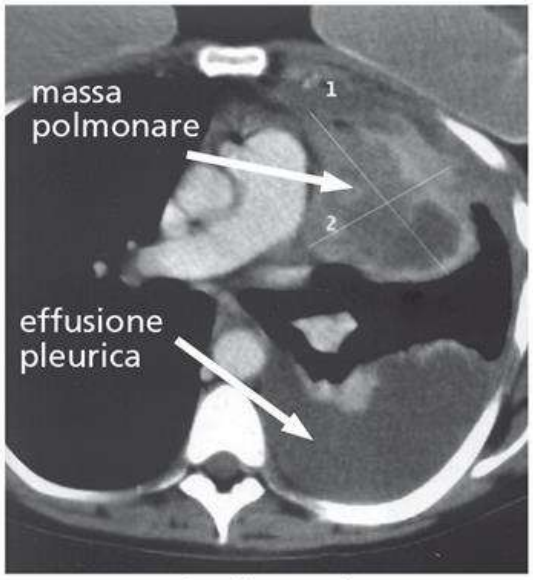
Nel 2018 venne pubblicato un altro studio, con il contributo, tra gli altri, di Rosenberg, che raccontava un nuovo successo clinico. Una donna di 49 anni affetta da un tumore al seno, allo stadio di metastasi, positivo per ER e negativo per HER2, che era refrattario a diversi processi di chemioterapia, venne coinvolta in un esperimento clinico, progettato per determinare l'abilità da parte dei TIL autologhi (cellule dello stesso paziente prelevate e reinfuse) di mediare la regressione del tumore in pazienti affetti da tumori epiteliali metastatici. Un sequenziamento di tutti gli esoni e dell'RNA di una lesione sottocutanea al seno destro rivelarono la presenza di 62 mutazioni somatiche non-sinonime. I TIL della paziente vennero resi reattivi alle versioni mutate di quattro proteine

(SLC3A2, KIAA0368, CADPS2 e CTSB). L'adozione di questo trasferimento di TIL specifici per proteine mutate in comunione con l'interleuchina 2 e il blocco al checkpoint hanno permesso di ottenere la completa regressione duratura del tumore al seno metastatico, regressione che stava durando da oltre 22 mesi, e rappresenta un nuovo approccio immunoterapeutico per la cura di questi pazienti.

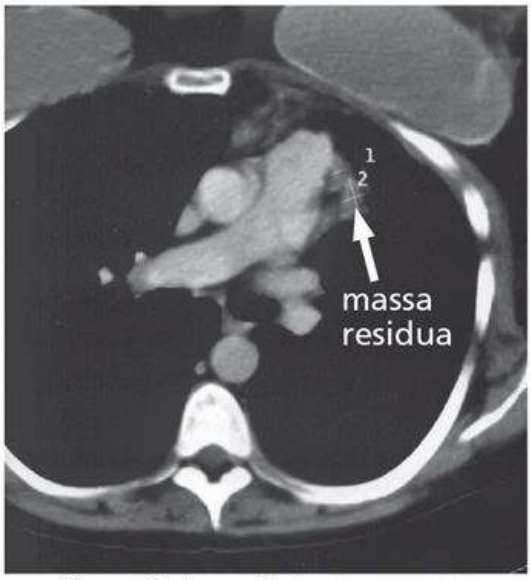
Nel 2018 il Premio Nobel per la Medicina è stato vinto da James P. Allison e Tasuku Honjo. I due scienziati avevano sviluppato, uno indipendente dall'altro, due trattamenti per la regressione dei tumori, dopo aver individuato due fattori chiave nella regolazione negativa della funzione immunitaria dei linfociti T: CTLA-4 (Cytotoxic T-Lymphocyte-associated Antigene 4) e PD-1 (Programmed cell Death protein-1).



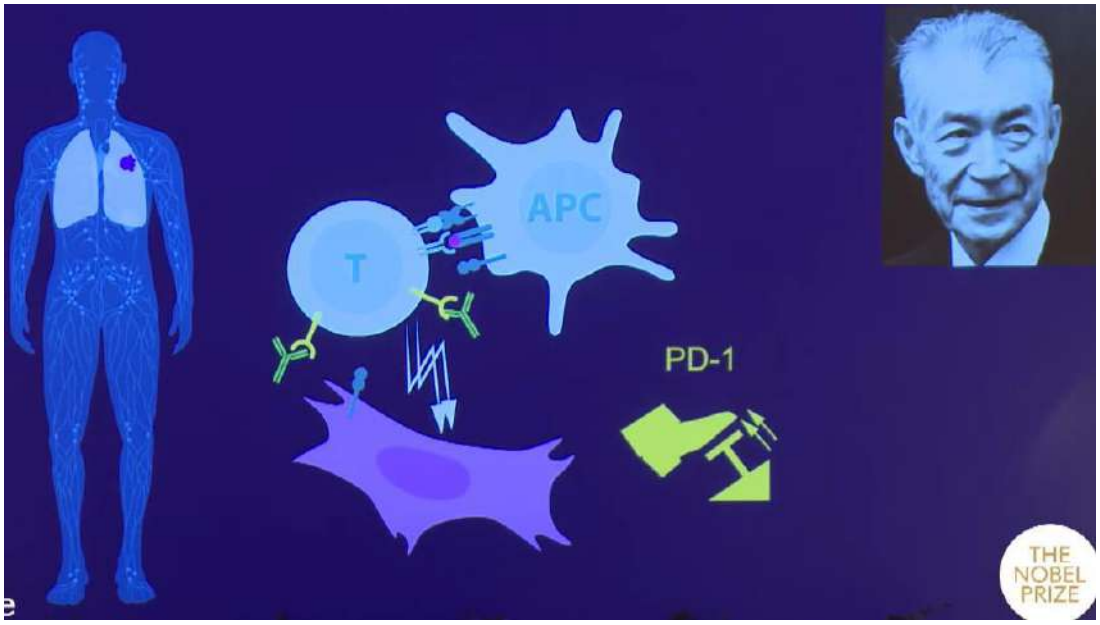


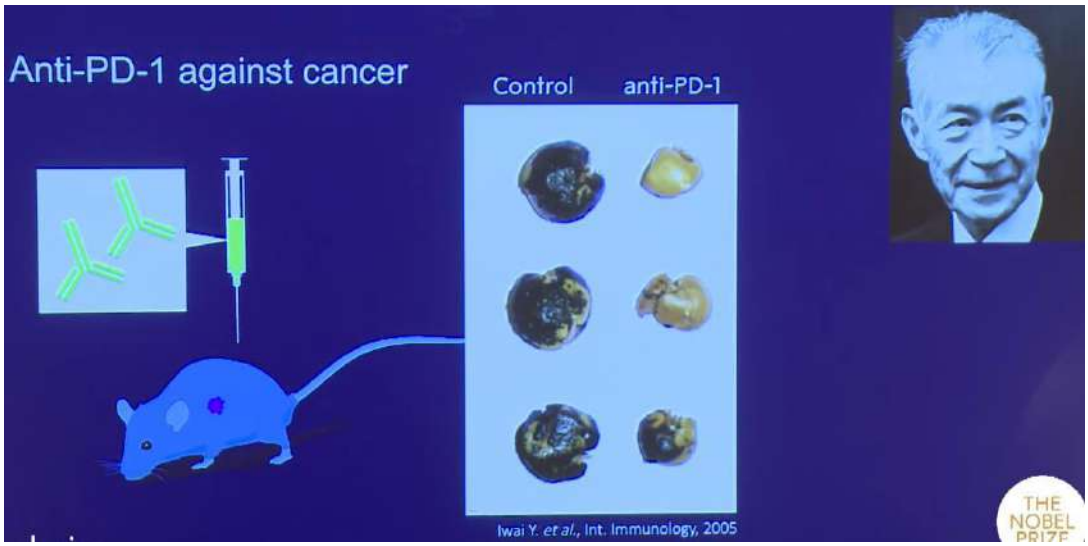


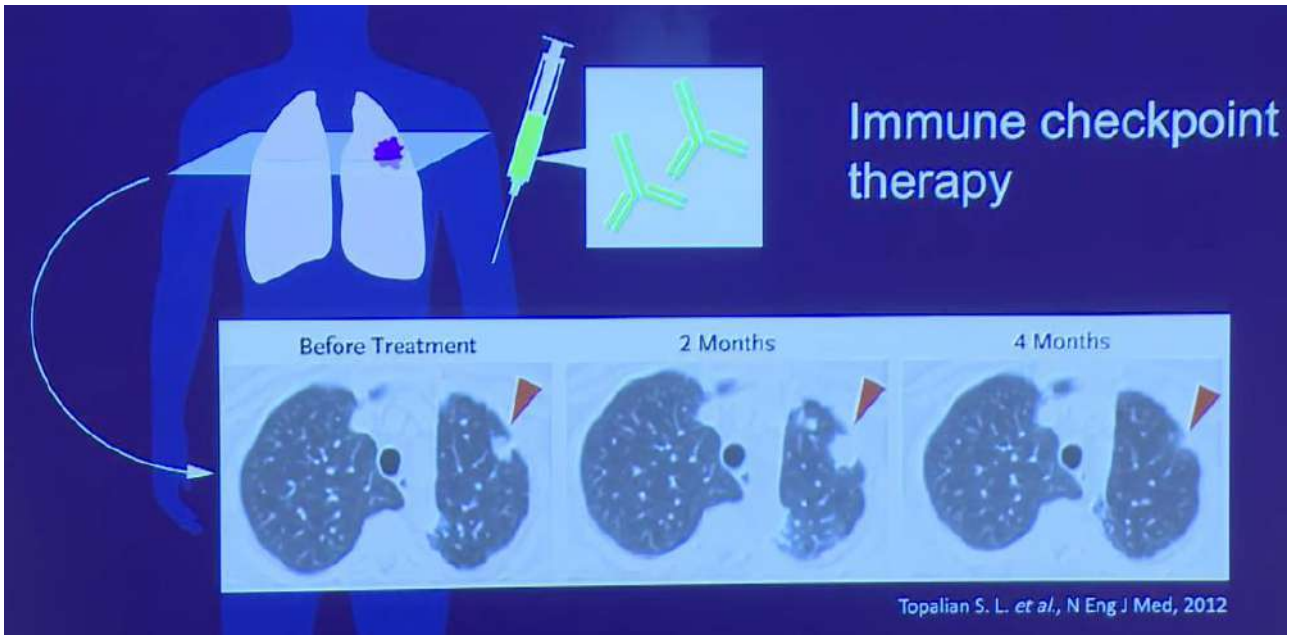
pretrattamento



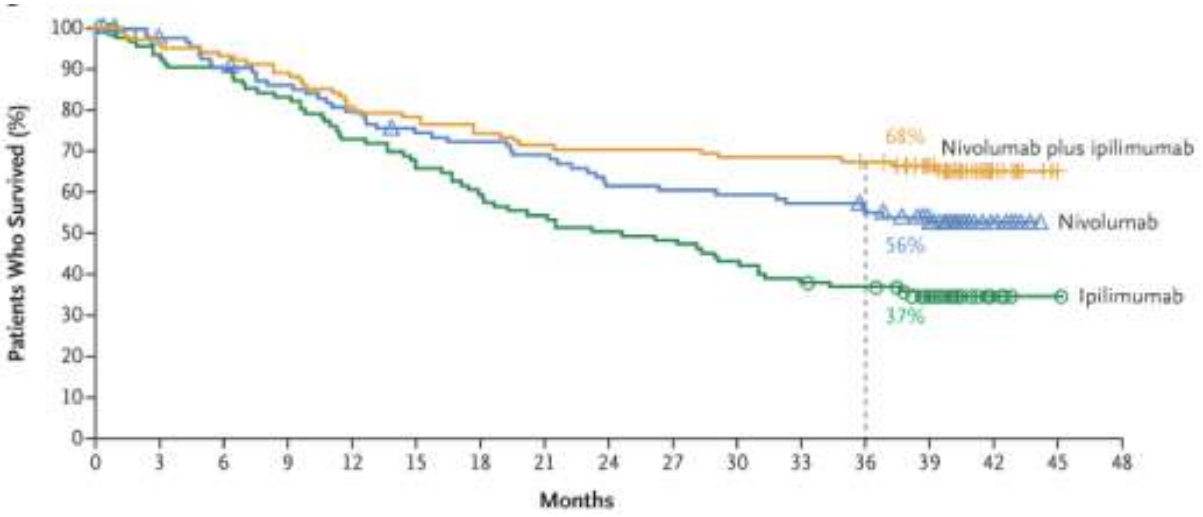
5 mesi dopo il trattamento





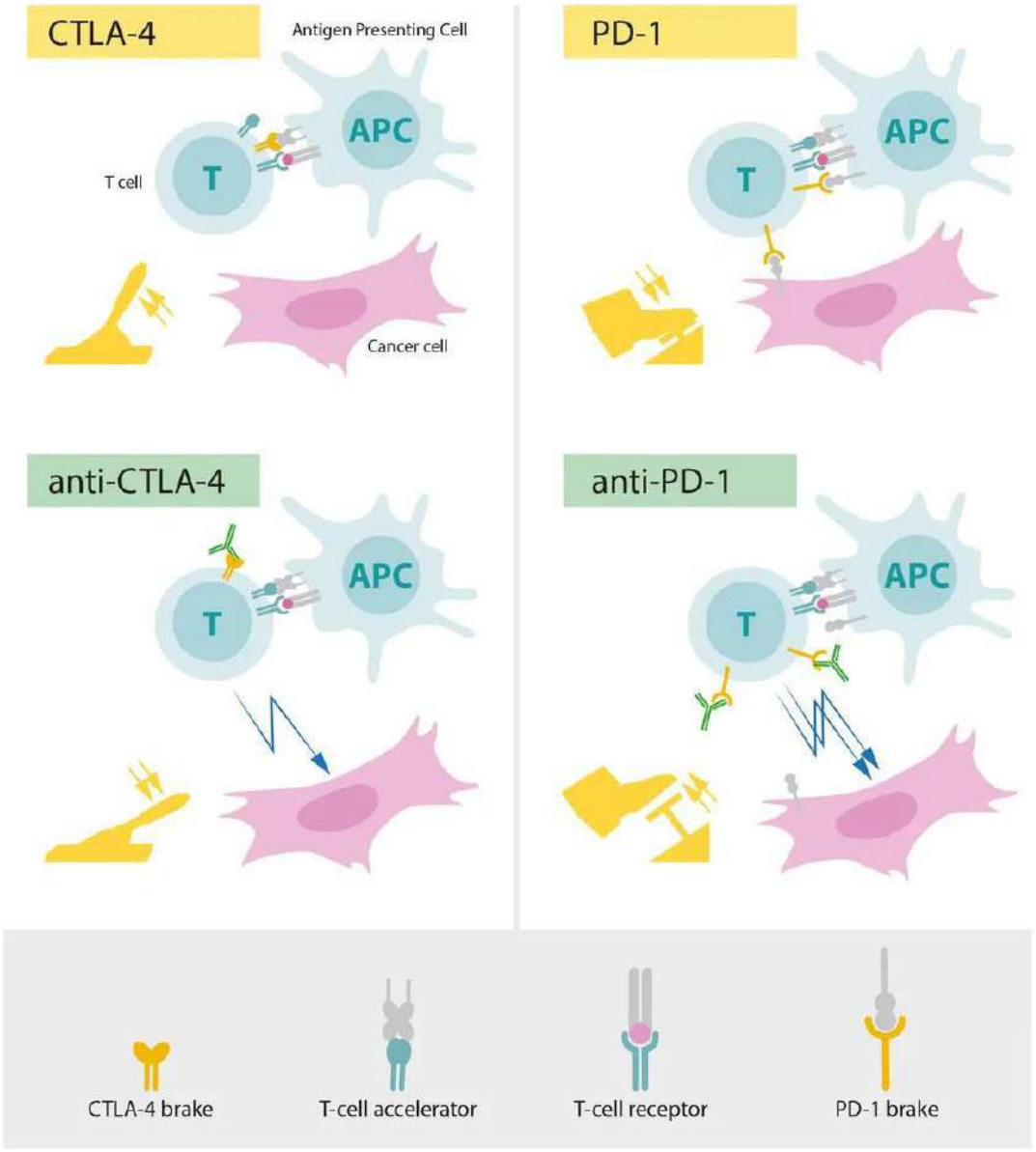


L'inibizione di questi target, che ha migliorato l'attivazione del sistema immunitario, ha condotto a nuove immunoterapie per il melanoma, per il tumore delle cellule dei polmoni e altri tipi di tumore. Ipilimumab, un inibitore di CTLA-4, è stato approvato per la cura di melanomi avanzati o non resecabili. Nivolumab e Pembrolizumab, entrambi inibitori di PD-1, sono stati approvati per curare pazienti affetti da un melanoma in stato avanzato o metastatico e pazienti affetti da un tumore polmonare refrattario e metastatico. In aggiunta, la combinazione di Ip e Niv è stata approvata in pazienti con un melanoma BRAF WT metastatico o non resecabile.



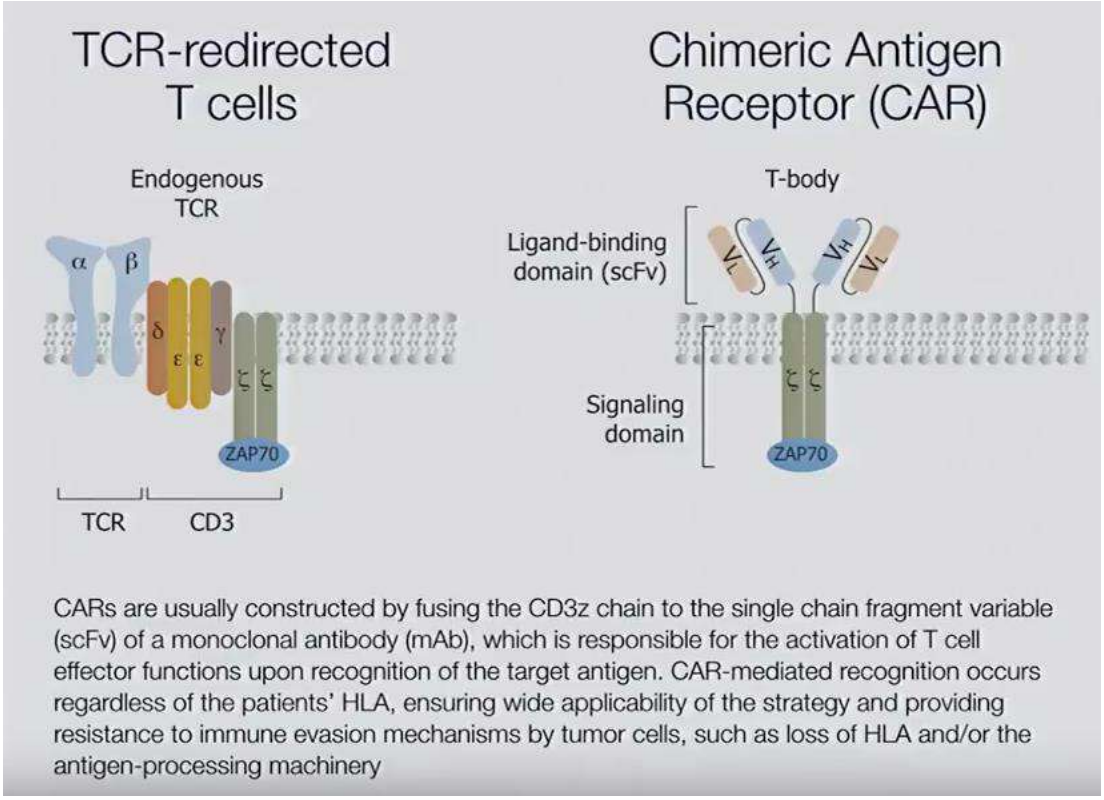
I ruoli di CTLA-4 e PD-1 nell'inibizione delle risposte immunitarie, incluse quelle antitumorali, sono ampiamente distinti. CTLA-4 si pensa regoli la proliferazione nelle prime fasi di una risposta immunitaria, in primo luogo nei linfonodi, mentre PD-1 sopprime i linfociti T nella fase tardiva della risposta immunitaria, in primo luogo nei tessuti periferici. I profili clinici degli agenti immuno-oncologici inibenti questi due 2 checkpoint differiscono proprio sulla base delle differenze a livello dei meccanismi che li contraddistinguono. CTLA-4, se unito ad un antigene artificiale, provoca una sorta

di “frenata” cellulare, che porta alla morte della cellula cancerosa, mentre il farmaco anti PD-1 attiva un’accelerazione del linfocita T, che uccide la cellula tumorale.

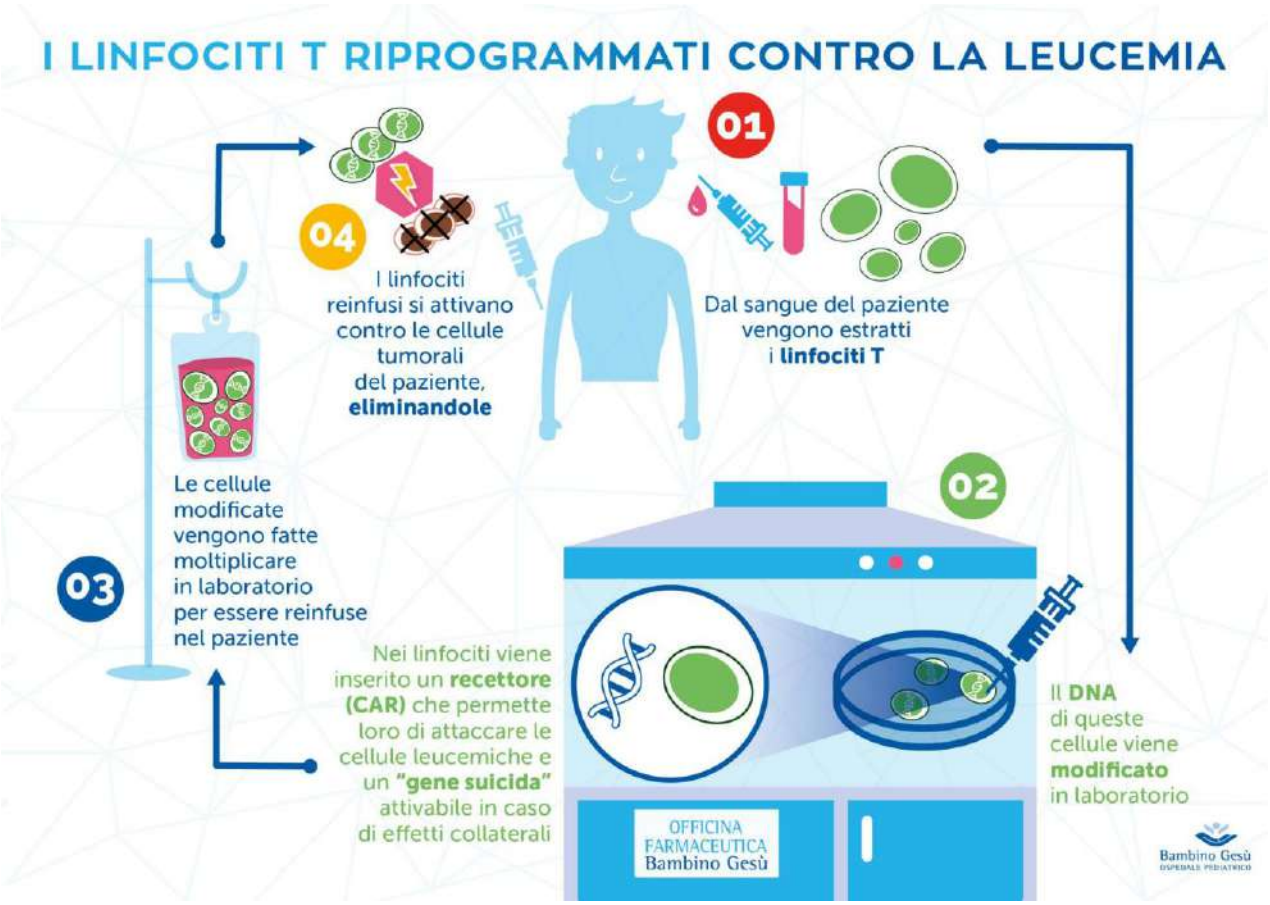


© The Nobel Committee for Physiology or Medicine. Illustrator: Mattias Karlén

Un’altra recente strategia per combattere la leucemia è stata messa in campo qualche anno fa: si tratta della terapia CAR-T. Si tratta di un’applicazione di ingegneria genetica ai linfociti T propri di ciascun paziente per rendere queste cellule più intelligenti e resistenti nello scovare e distruggere il cancro. Le cellule T, quindi, sono “personalizzate” per concentrarsi sul tipo specifico di cancro da cui è affetto il paziente. Comincia tutto quando il sangue del paziente viene filtrato in un macchinario che separa le cellule T e altri globuli bianchi dal resto. Gli scienziati mescolano le cellule T con un virus che è stato disattivato, così non provocherà malattie. Invece, quel virus trasporta istruzioni genetiche affinché le cellule T producano un recettore artificiale chiamato “recettore chimerico per l’antigene” (CAR). Esso rintraccia il bersaglio canceroso e si attiva per attaccarlo. Milioni di copie di queste cellule ingegnerizzate vengono fatte crescere in laboratorio e reinfuse nel paziente in endovena. Le



prime cellule CAR-T sono state testate contro forme di leucemia e linfomi, che portano un marker o un antigene dal nome CD19. Una volta nel torrente sanguigno le cellule CAR-T riconoscono quell'antigene e si attaccano. Rilasciano sostanze chimiche tossiche che innescano la morte della cellula cancerosa. Non appena una cellula cancerosa muore le cellule T-CAR proseguono alla successiva, moltiplicandosi nel sangue del paziente, finchè non è più necessario il loro intervento. Vengono allora spente attraverso un "interruttore" genico, che le porta a "suicidarsi".

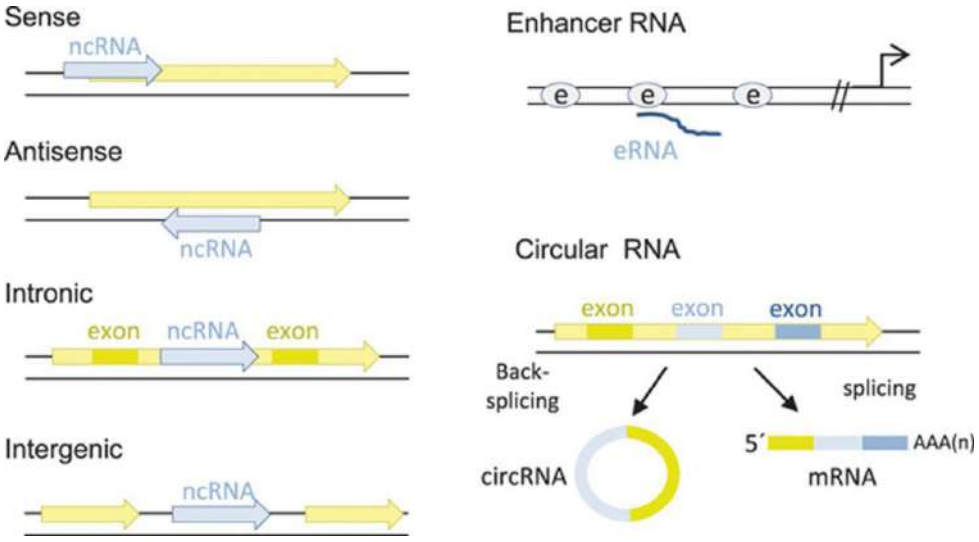


Le nuove terapie in corso di sviluppo e quelle del futuro saranno concentrate sui checkpoint del sistema immunitario.

I long non-coding RNA

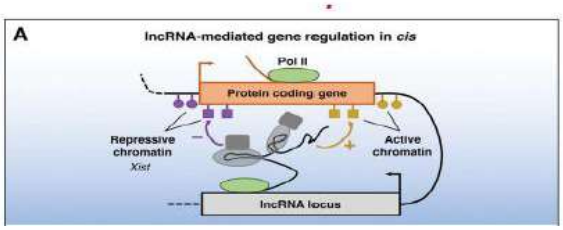
La rivoluzione dell'RNA

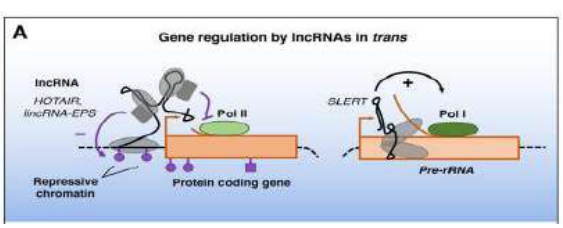
Ci sono circa 20 000 che codificano proteine e si è visto con l'aumentare della complessità degli organismi aumentava la porzione di DNA non codificante. Con le tecniche di sequenziamento si è capito che i trascritti non proteici erano importanti. L'80% del DNA, infatti, viene trascritto. I microRNA, la cui lunghezza è inferiore ai 200 nt, sono trascritti dall'RNA polimerasi II e vengono processati da endonucleasi, mentre i lncRNA si riconoscono come RNA più lunghi (>200 nt), sono processati sempre dall'RNA polimerasi II, possiedono un cap al 5' e una coda parzialmente poliadenilata, sono soggetti a splicing e non si sono molto conservati nell'evoluzione. Sono in grado di svolgere molte e diverse funzioni. Vengono classificati in base alla loro posizione:

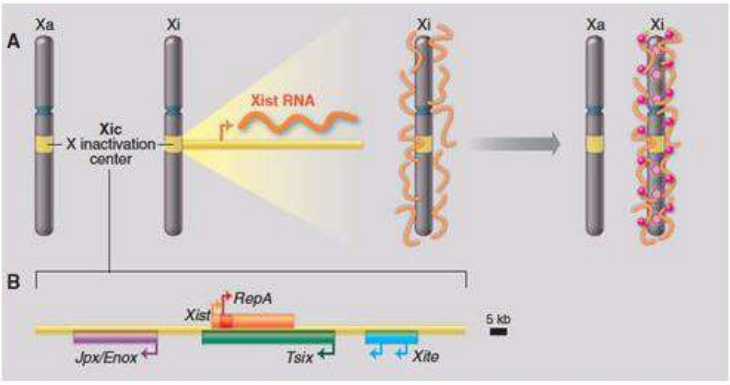


- Senso;
- Antisenso;
- Intronici;
- Intergenici;
- RNA enhancer;
- RNA circolare.

Possono anche essere distinti tra cis- e translncRNA. I cis- influiscono sull'espressione e sullo stato della cromatina dei geni vicini, mentre i trans- riescono a raggiungere porzioni di genoma anche distanti dalla loro posizione.



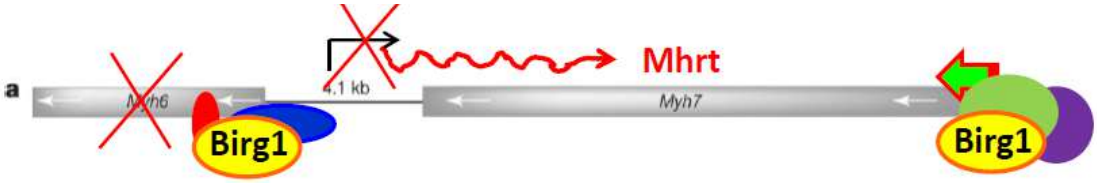




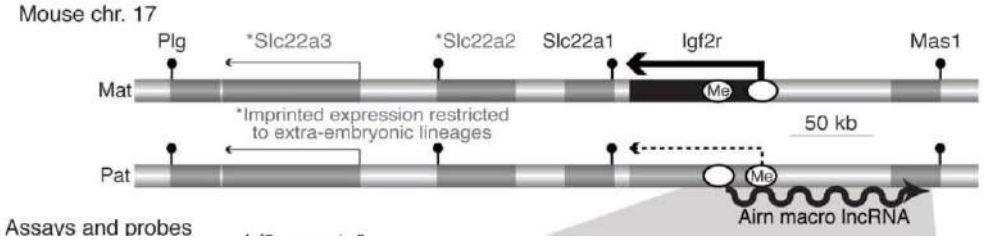
Un lncRNA in cis- è XIST, il primo a essere stato descritto dal punto di vista funzionale. E' coinvolto nel processo di inattivazione del cromosoma X, necessario per compensare il dosaggio genico nelle donne rispetto agli uomini. XIST, naturalmente, è espresso in misura differente in un sesso rispetto all'altro. Il lncRNA è abbondante e inizia ad interagire con i repressori del cromosoma X, provocando l'ultracondensazione di X, a formare il cosiddetto corpo di Barr, intrascrivibile, che rimarrà tale per tutto il resto della vita della femmina. A questo punto Xist esaurisce la propria funzione e viene degradato. E' localizzato sul cromosoma X nel centro di inattivazione dell'X (Xic).

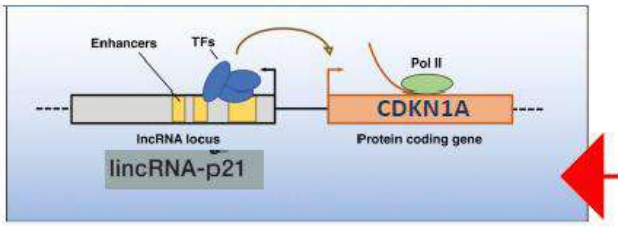
Un altro esempio è il MYHEART, Myosin Heavy-Chain-Associated RNA Transcript. La catena pesante beta della miosina è tipica della fase fetale per la contrazione lenta, ma, successivamente, viene rimpiazzata dalla catena alfa (che fornisce più potenza alla miosina). Questi due geni sono molto conservati. Questi due geni sono molto conservati. Il MYHEART va ad agire in seguito all'infarto.

Un tessuto cardiaco, dopo l'infarto, presenta aberrazioni nell'espressione genica e una miopatia che blocca la sintesi (attivando la proteina elicasi) di alfa-miosina e l'espressione di MYHEART e stimola quella di beta-miosina. Questo fenomeno provoca ipertrofia del tessuto cardiaco. Se invece si fornisce MYHEART al cuore, il lncRNA attacca il dominio dell'elicasi di Birg1 per evitare che esso condensi la cromatina del gene dell'alfa-miosina, riconosciuta proprio attraverso il dominio elicasico.

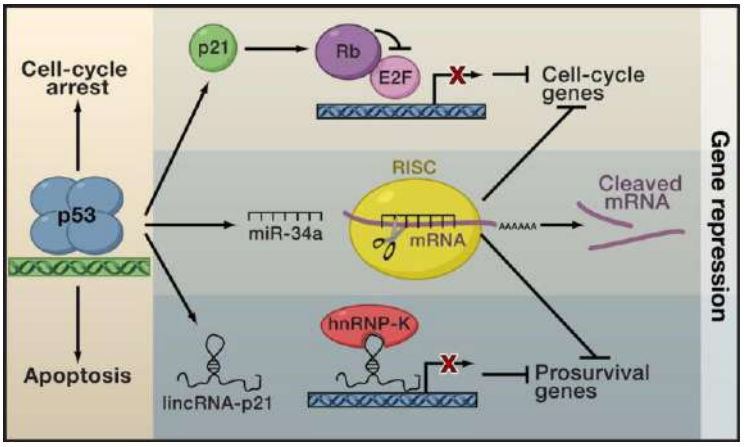


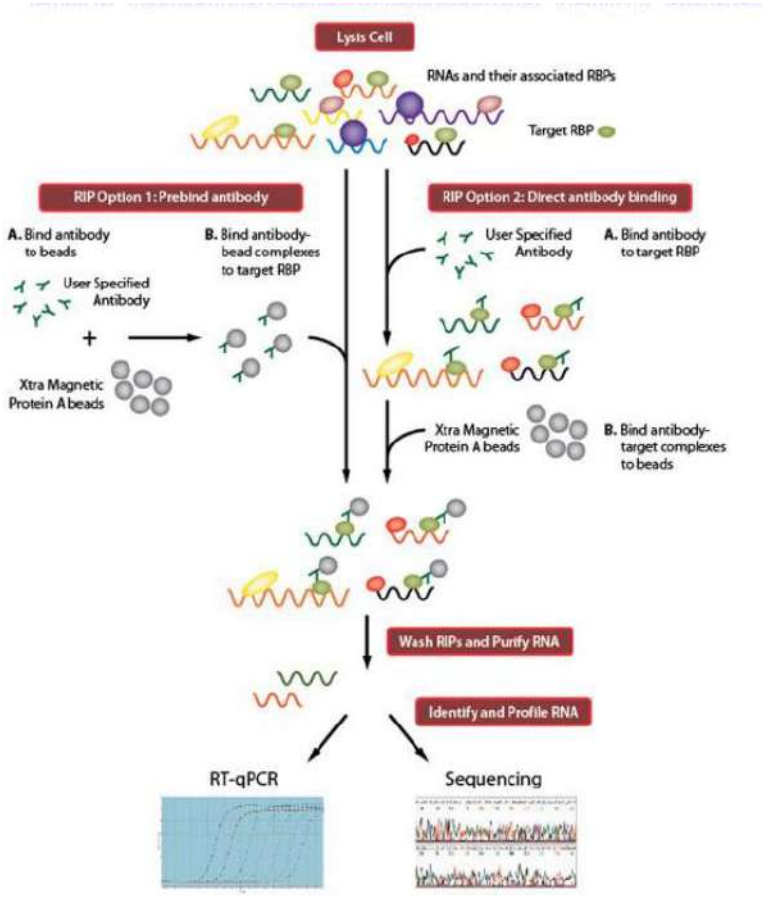
Il gene Igf2r è soggetto a imprinting materno e in direzione antisenso trascrive un non-coding, che è in grado di reprimere Ig2fr coprendone il promotore, impedendo la trascrizione del gene e interferendo con il reclutamento dell'RNA polimerasi II in assenza di cromatina repressiva. Il lncRNA che disegna questa "nuvola" è Airn RNA.





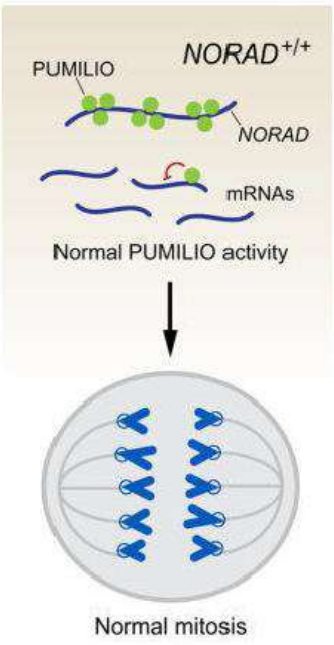
Il lincRNA-21 è un lncRNA dipendente da p53 (soppressore tumorale), che attiva p21 (gene target) e blocca la trascrizione dei geni della proliferazione cellulare. Il lincRNA p21 è prodotto da una sequenza intergenica che è vicina del gene CDKN1A, che codifica per p21. La sua sequenza è importante perché contiene gli enhancer di p21 e lega diversi fattori di trascrizione coinvolti nella trascrizione di CDKN1A. P21 è importante, a sua volta, perché attiva i geni del ciclo cellulare, che permettono alla cella di dividersi, e abbassa i livelli di p53, inducendo il taglio dell'mRNA di quello. Svolge dunque anche un ruolo modulatorio sull'espressione di CDKN1A.



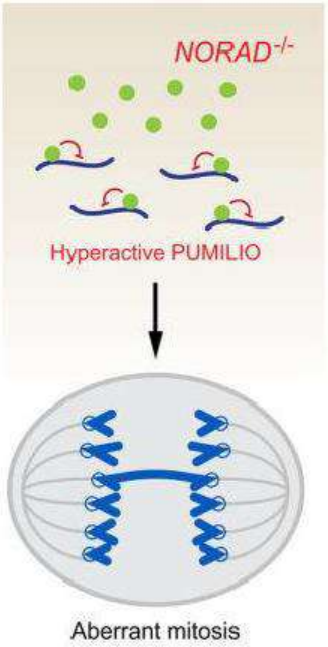


I trans lncRNA, come già detto, interferiscono con geni lontani e interagiscono con i loro trascritti. In un lavoro del 2012 Guttman e Lander si erano chiesti qual era la funzione dei trans lncRNA nelle cellule embrionali. Infettarono le cellule con shRNA (RNA a forcina corta), che si legarono ai non-coding, che si degradavano successivamente. Secondo lo screening effettuato, il silenziamento di più del 90% del lincRNA provocava importanti cambiamenti nell'espressione genica, che in parte erano in quantità uguale ai cambiamenti di espressione causati dallo spegnimento di geni codificanti per proteine regolatrici. Hanno dunque un grande impatto sullo sviluppo. Sono importanti per la staminalità (NANOG/OCT4/SOX). In seguito, il team di Guttman procedette ad un saggio di immunoprecipitazione dell'RNA (RIA), nel quale viene trascinato in basso l'RNA genomico legato. L'RNA precipitato venne analizzato, servendosi di anticorpi contro i complessi cromatinici del 12. Gli autori identificarono 74 lincRNA che erano legati

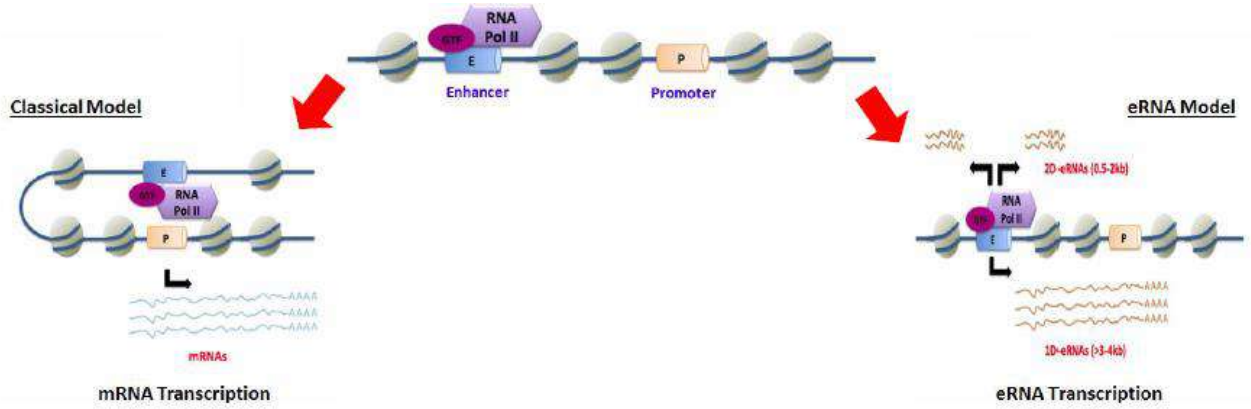
a più di un complesso cromatinico. In linea con i risultati ottenuti, fu proposto un modello nel quale i lincRNA funzionano come impalcature flessibili specifiche per tipo cellulare per reclutare complessi proteici, come i fattori epigenetici, per avviare programmi di trascrizione.



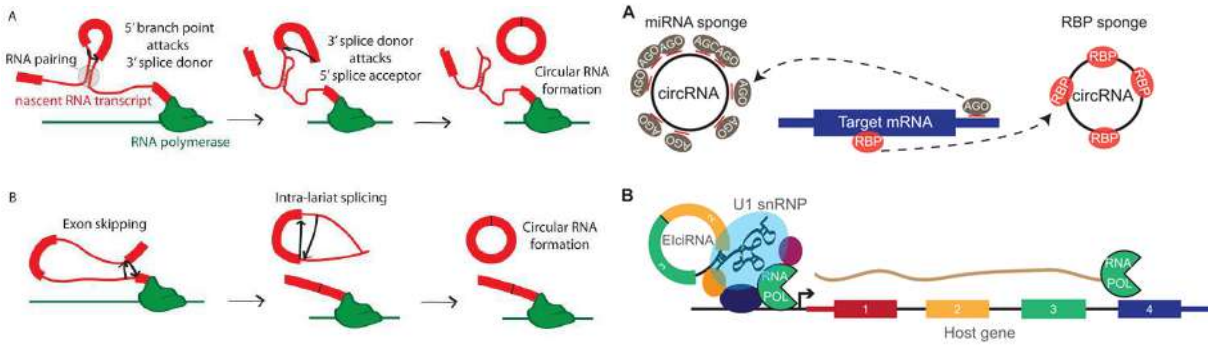
NORAD è un lncRNA indotto in seguito a danni al DNA. La sua inattivazione innesca una drammatica aneuploidia in linee cellulari precedentemente stabili dal punto di vista del cariotipo. NORAD mantiene la stabilità genomica sequestrando le proteine PUMILIO, che reprimono la stabilità e la traduzione degli RNA messaggeri a cui si legano. In assenza di NORAD, le PUMILIO provocano instabilità cromosomica reprimendo iperattivamente fattori della mitosi, di riparazione e replicazione del DNA.



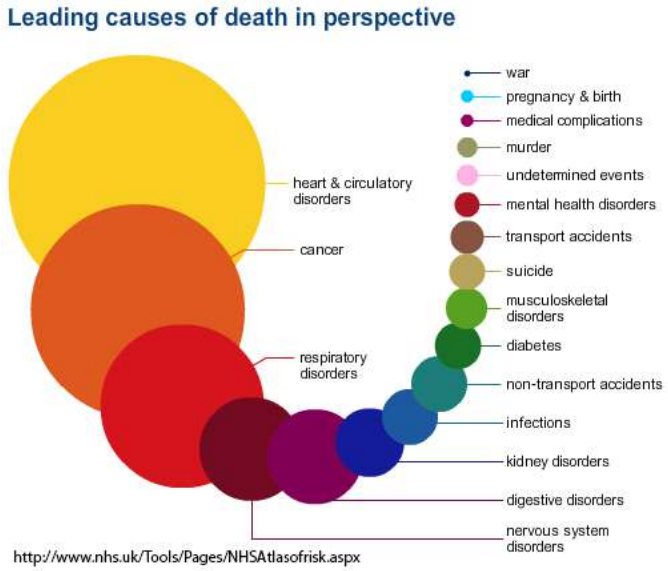
Perché avvenga la trascrizione, la cromatina deve aprirsi per permettere il legame dei macchinari di trascrizione (ad esempio, la RNA polimerasi II) con il sito di trascrizione. Questo processo inizia con il legame dei fattori di trascrizione al sito detto enhancer. La RNA Pol II, in base all'attività da svolgere, si lega a migliaia di enhancers. In corrispondenza di questi punti trascrive bidirezionalmente una nuova classe di lncRNA (gli eRNA), entro il dominio dell'enhancer. Sono meno espressi degli altri lncRNA e non subiscono splicing.



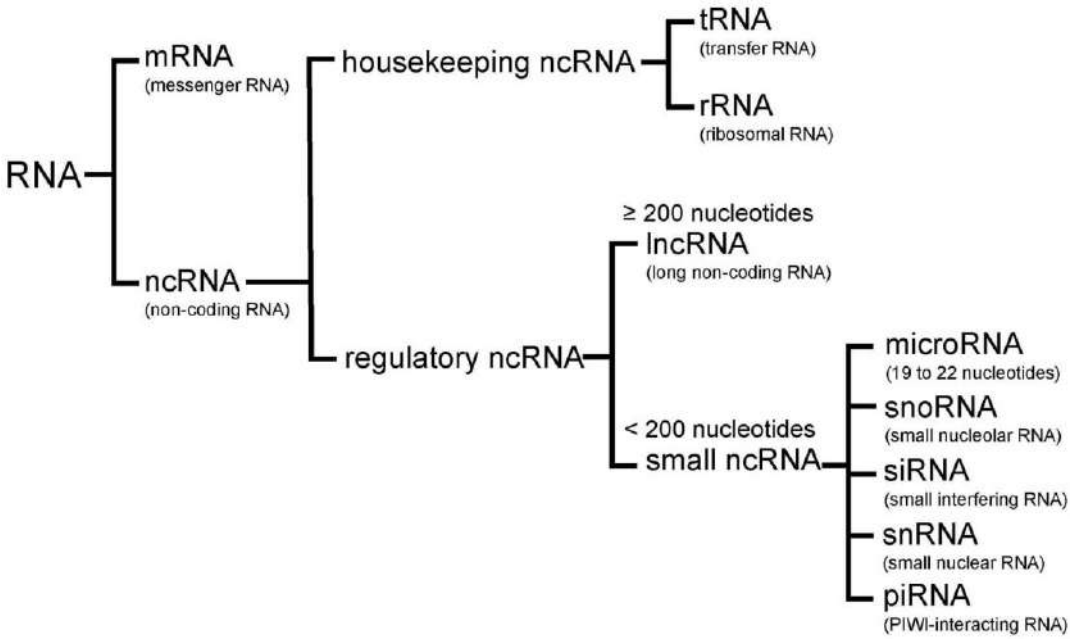
I circRNA aumentano la frequenza dello splicing alternativo nell'uomo, derivano dallo splicing di un trascritto (da introni o da esoni) e assumono diversi comportamenti. Ad esempio, possono funzionare come spugne, contenenti diversi siti di legame per particolari miRNA o RBP, che vengono così sottratti all'mRNA bersaglio, provocando un'alterazione dell'espressione genica. Oppure interagiscono con la snRNP U1 e con altri complessi trascrizionali, inducendo la trascrizione dei geni ospiti. Spesso si originano da fenomeni di backsplicing, dopo un exon skipping.



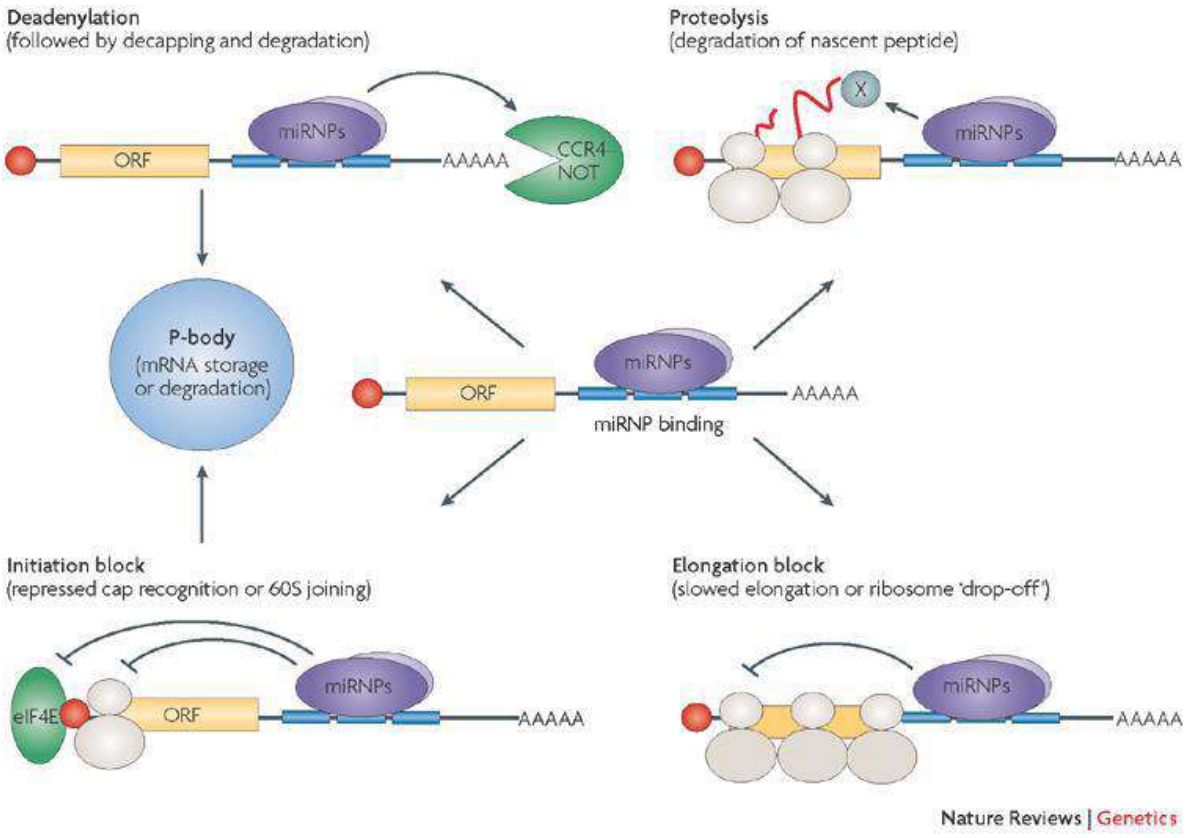
Gli pseudogeni nel cancro



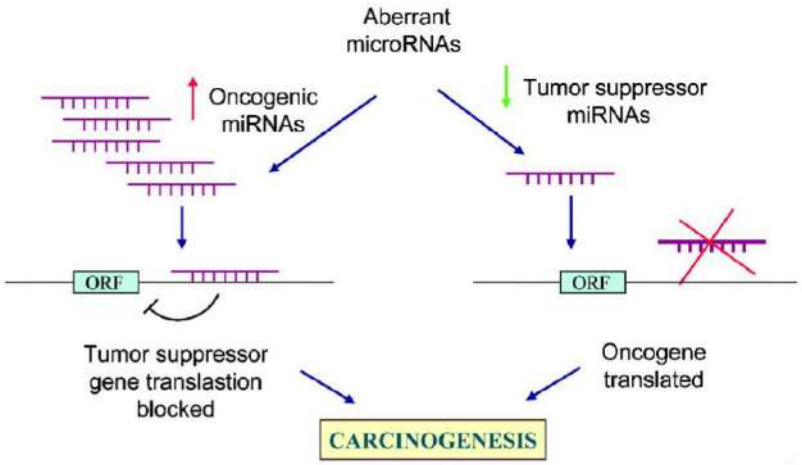
Il cancro è la seconda causa di morte nei Paesi sviluppati, dopo le malattie cardiovascolari. Si definisce cancro una crescita cellulare anormale che ha la possibilità di diffondersi in altri distretti corporei. I geni coinvolti nel cancro si dividono in due categorie: gli oncogeni e gli oncosoppressori. Gli oncogeni sono geni che producono proteine capaci di trasformare cellule in coltura o di indurre il cancro negli animali (acceleratori). Il proto-oncogene BRAF, ad esempio, in seguito a mutazione puntiforme, ad amplificazione genica o a traslocazione cromosomica riceve un gain-of-function, quindi si attiva e diventa un oncogene (BRAFV600E), che emette un segnale incontrollato di crescita cellulare. I geni oncosoppressori, invece, inibiscono la proliferazione cellulare (freni). Un famoso oncosoppressore è p53, che, in seguito a mutazioni puntiformi, a delezione, a silenziamento epigenetico, subisce una loss-of-function, diventando una proteina mutante, non più funzionale.



I geni del cancro non sono soltanto codificanti (cioè trascrivono un RNA messaggero), ma possono anche non dare un prodotto proteico, sintetizzando molecole di RNA non messaggero. I non-coding RNA possono essere "housekeeping"(rRNA e tRNA) oppure con funzione regolatoria. Questi ultimi si dividono in lncRNA e sncRNA (lunghezza inferiore ai 200 nt). I sncRNA contengono, all'interno della loro classificazione, i miRNA, piccole molecole di RNA con lunghezza compresa tra 19 e 22 nt. Si legano alla regione 3' UTR degli mRNA e ne reprimono la traduzione. Gli RNA messaggeri, dunque, vanno incontro a degradazione attraverso varie modalità: deadenilazione, proteolisi, blocco iniziale, blocco in fase di allungamento. Il degradamento degli mRNA è responsabile dell'84% delle diminuzioni osservate di proteine prodotte.



I MRE sono gli elementi di riconoscimento dei microRNA, tra i quali c'è la regione seed, l'appaiamento centrale non perfetto, la conservazione, la posizione e il contesto. Se i miRNA sono aberranti, si verificano due tipi di situazioni legate al cancro. Si può osservare un aumento dei miRNA che si legano agli oncosoppressori, cioè di miRNA oncogenici, bloccandone la traduzione.



Analogamente, se diminuisce il numero di miRNA oncosoppressori la traduzione degli oncogeni si intensifica. In entrambi i casi il risultato è la carcinogenesi.

I NAT sono trascritti che coinvolgono entrambi i filamenti, sia quello senso che quello antisenso. Se entrambi i filamenti sono trascritti dalla stessa unità si parla di cisNAT, altrimenti, quando si attivano due unità trascrizionali, si parla di transNAT. I lncRNA sono simili a proteine: sono regolatori epigenetici, trascrizionali post-trascrizionali.

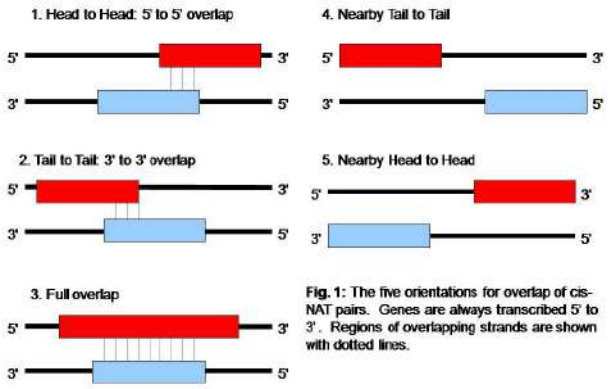
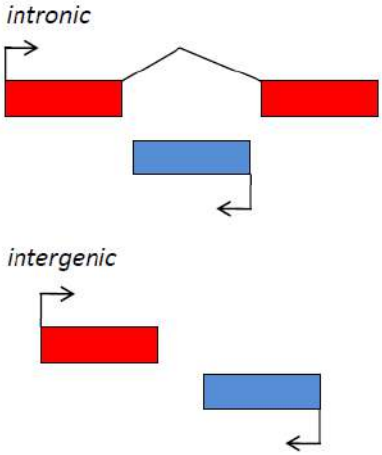
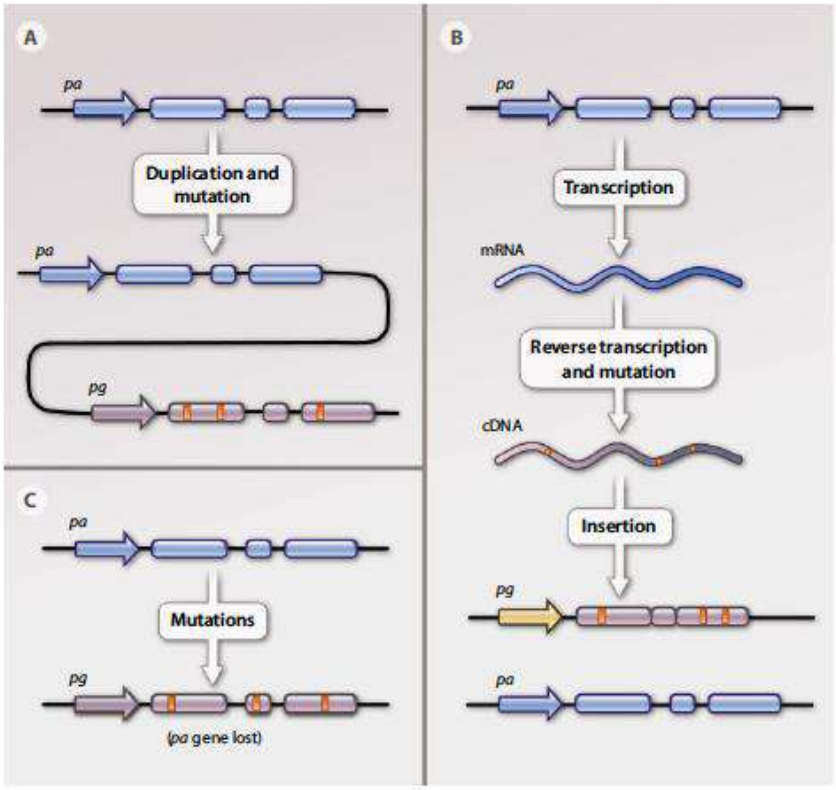


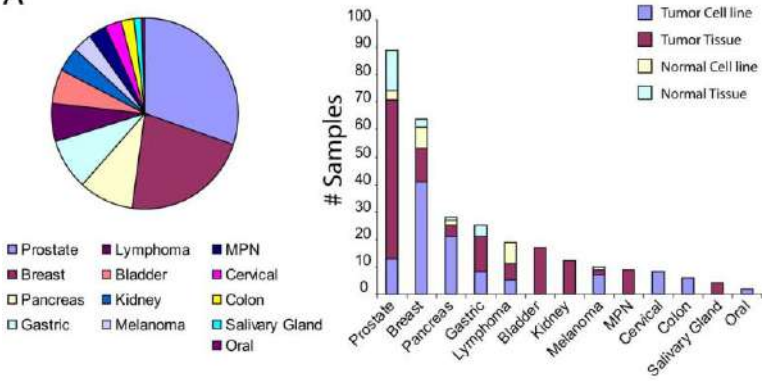
Fig. 1: The five orientations for overlap of cis-NAT pairs. Genes are always transcribed 5' to 3'. Regions of overlapping strands are shown with dotted lines.





Gli pseudogeni sono elementi genetici strutturalmente simili a geni che codificano proteine funzionali, ma contengono "difetti" che li rendono incapaci di codificare proteine che siano pienamente funzionali. Un tipo di pseudogeni, tipo A, si originano da duplicazione e conseguente mutazione. Si chiamano pseudogeni non processati. Può essere uno dei due geni parentali. Si trova quindi sullo stesso cromosoma del gene funzionante, ne condivide i promotori e contiene introni ed esoni. Il tipo B si origina da trascrizione, trascrizione inversa e mutazione. Si inserisce, infine, con un "landing", in un'altra porzione del genoma. L'ultimo tipo, C, è costituito dagli pseudogeni unitari: geni un tempo funzionanti che hanno accumulato nel tempo mutazioni che hanno causato una loss of function anche per il ruolo dell'RNA.

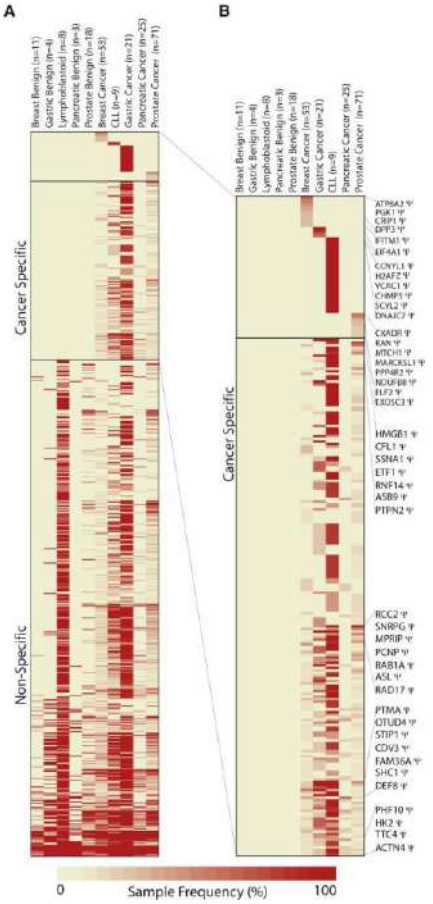
A



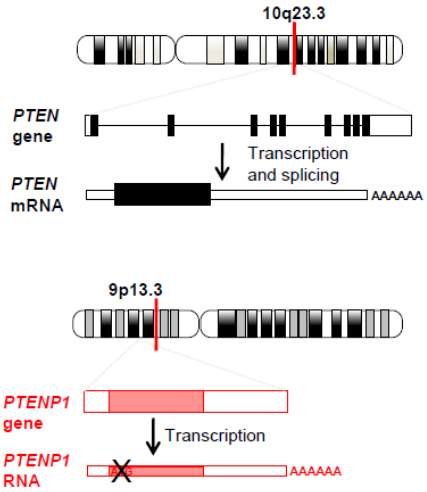
B

Tissue Type (13)	Tumor		Benign		Total Sample	Total PF Read (million)	PF Read average (million)
	Cell line	Tissue	Cell line	Tissue			
Prostate	13	58	3	15	89	1492.0	16.8
Breast	41	3	8	12	64	653.1	10.2
Pancreas	21	4	2	1	28	357.8	12.8
Gastric	8	13		4	25	351.3	14.1
B-Lymphoid	5	6	8		19	383.5	20.2
Bladder		17			17	203.9	12.0
Kidney		12			12	190.3	15.9
Melanoma	7	2		1	10	123.4	12.3
MPN		9			9	233.2	25.9
Cervical	8				8	140.5	17.6
Colon	6				6	161.8	27.0
Salivary Gland		4			4	103.7	25.9
Oral	2				2	56.9	28.4
Total					293	4451.3	15.2

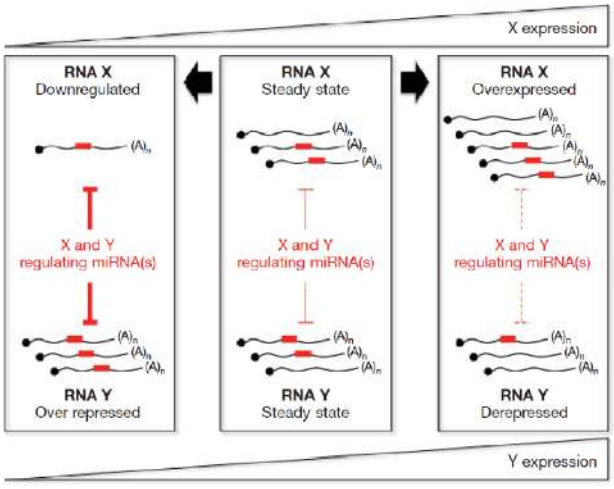
Nel genoma usano esistono circa 18 000 pseudogeni, che codificano per circa 100 proteine. Si è osservato che l'espressione dei PG non è casuale nell'insorgenza di tumori. Alcuni PG



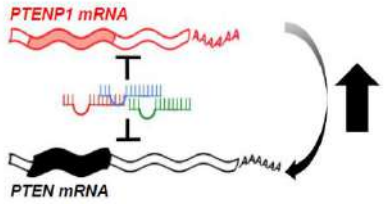
vengono espressi in casi di tumore, altri mai. Altri ancora si attivano soltanto in tipi specifici di tumore. Abbiamo PG cancer-specific, cancer type-specific e cancer subtype-specific.



Possono avvenire mutazioni a carico dei PG, come nel caso di PTENP1 (PG di PTEN), che viene deleto in modo focale, cioè vantaggioso per il tumore, nel cancro al colon. E' composto da 3 geni, il secondo dei quali viene colpito dalla mutazione "driver", mentre le successive



vengono considerate "passenger". L'assenza di trascrizione di PTENP1 favorisce lo sviluppo del tumore, in particolare se è dovuta alla presenza di agenti mutageni che, sotto forma di cDNA, alterano l'espressione dei geni ai quali si legano. Dagli pseudogeni si trascrive un mRNA e questo RNA aiuta quello dei geni parentali.



Entrambi i messaggeri, infatti, sono spugne per i miRNA. PTENP1 compete con PTEN per legare i miRNA. Se sono presenti più molecole di mRNA dello pseudogene, i miRNA tendono ad associarsi a quella "spugna", che ne assorbe di più, favorendo la derepressione del gene parentale.

Se c'è meno mRNA dello pseudogene (si dice che è downregulated), miRNA si accumuleranno sul gene funzionale, causandone l'overrepressione. Se invece la mutazione è sulla tripletta AUG iniziale, non avviene nessuna traduzione. Se la mutazione è di altro tipo, si riflette sulla proteina abnorme.

I PG sono coinvolti anche in processi di regolazione epigenetica, in particolare di silenziamento. OCT4 è una proteina espressa nelle cellule staminali. OCT4-pg4 è un PG capace di indurre il silenziamento trascrizionale di OCT4 reclutando un complesso di silenziamento multiproteico che si lega al promotore della proteina. La concentrazione di OCT4 cellulare si riduce e permette il processo di differenziamento cellulare.

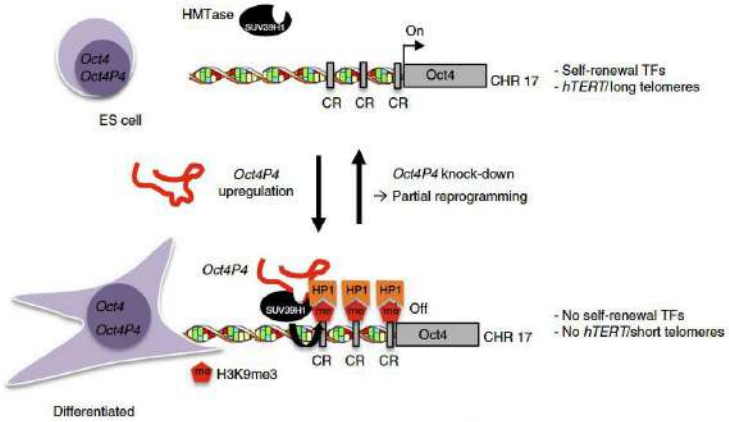
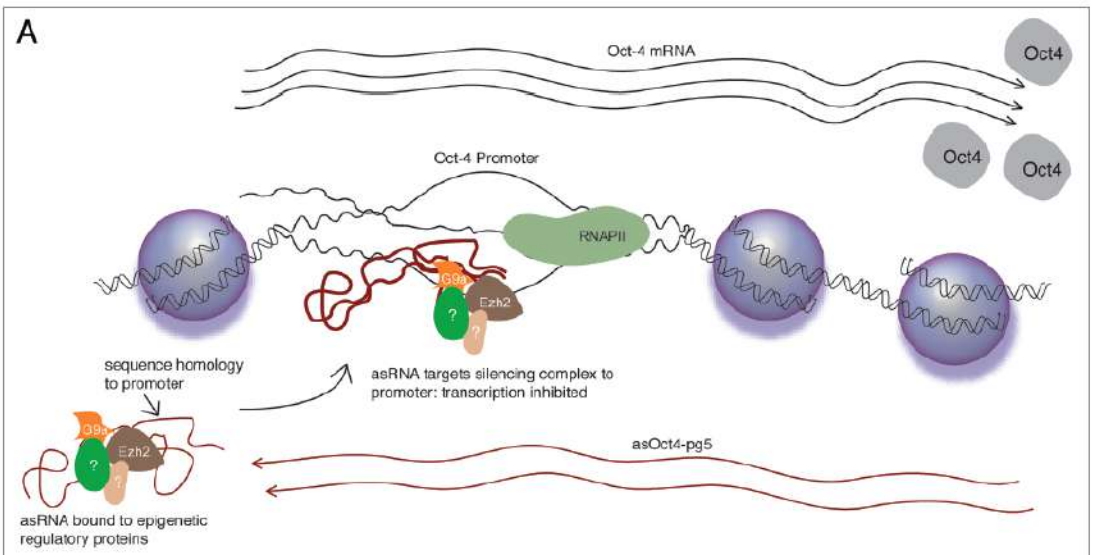
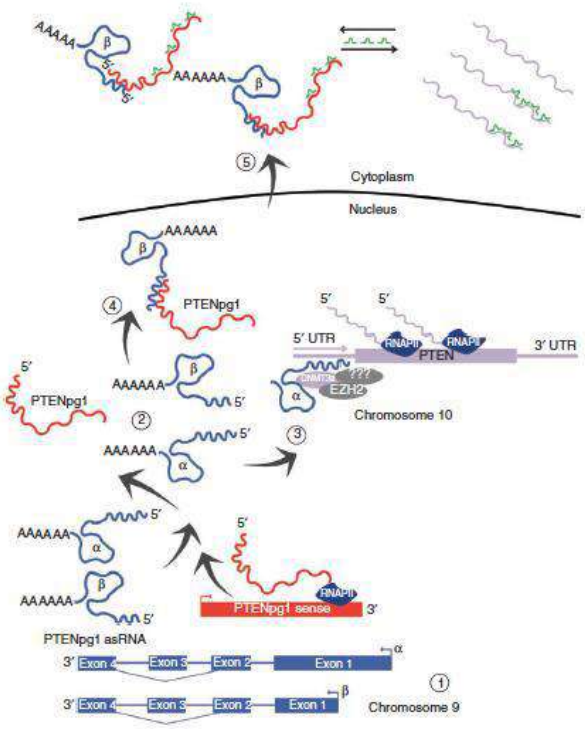


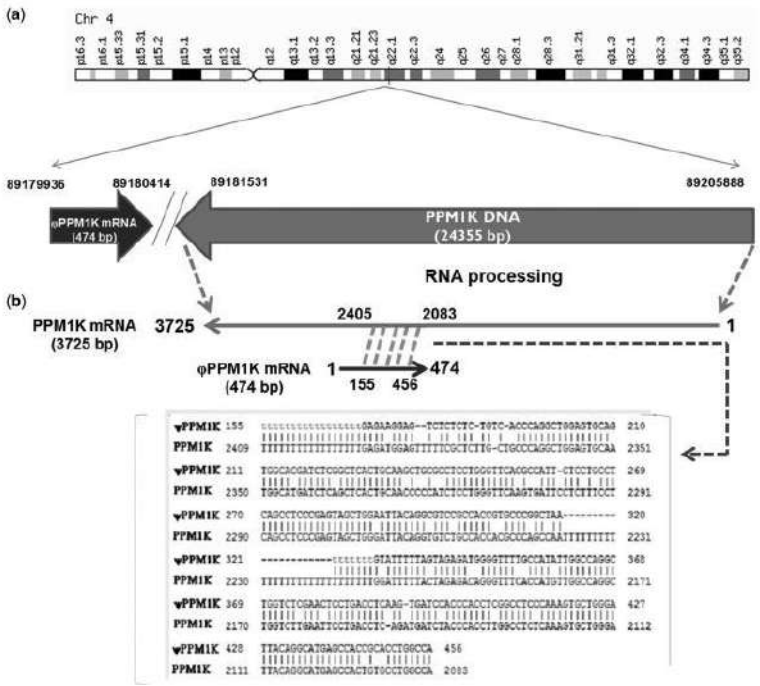
Figure 6 | A model for *Oct4P4* lncRNA function. The nuclear, long non-coding *Oct4P4* is upregulated during mESC differentiation and recruits the H3K9-specific HMTase SUV39H1 to impose H3K9me3 and HP1α at the promoter of ancestral *Oct4* gene in *trans*, leading to *Oct4* gene silencing. Depletion of *Oct4P4* lncRNA in differentiated pMEFs results in re-activation of *Oct4* expression and the acquisition of classic features of self-renewal such as increased expression of self-renewal transcription factors and activation of telomerase-dependent telomere maintenance mechanisms. This indicates that *Oct4P4* prevents the aberrant expression of self-renewal programmes in differentiated cells.



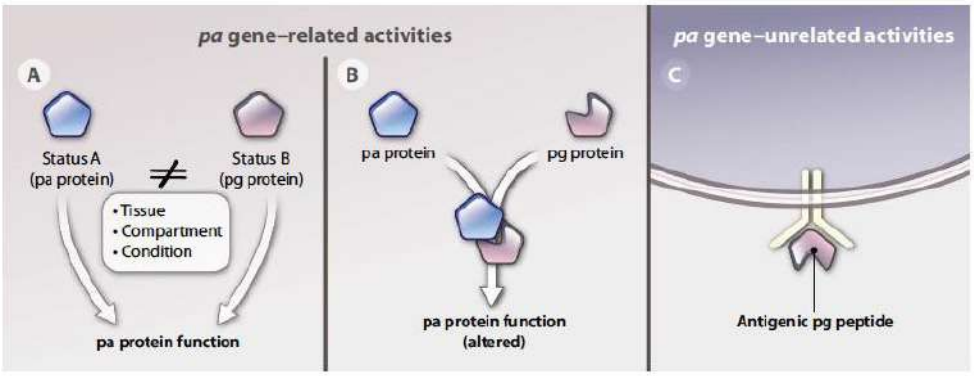


Per i PG il filamento senso stabilizza quello antisenso. L'antisenso beta aiuta la trascrizione di PTENP1, mentre l'antisenso alfa fa l'opposto. I loro livelli sono precisamente regolati. Quali sono, però, i rapporti di forza relativi? Chi "vince"?

Esistono anche gli endosiRNA, RNA a doppio filamento, che, a causa proprio della loro inusuale forma double-stranded, sono percepiti dalla cellula come invasori. Un esempio è lo pseudogene di PPM1K che viene trascritto, in una sequenza sovrapposta a quella del gene PPM1k, seguendo il filamento antisenso. Lo pseudogene in questione produce un double-stranded RNA con l'RNA prodotto dal gene funzionale, che può essere processato per diventare un endosiRNA



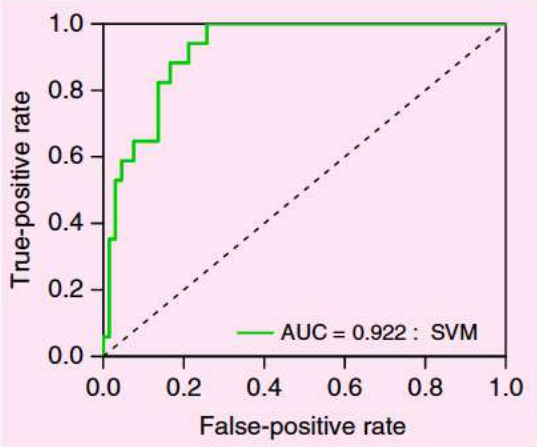
Tra le proteine wild-type provenienti da PG ricordiamo CYT2D7, espressa nel cervello. Il gene NANOG è responsabile, invece, della staminalità delle cellule, mentre NANOGP8 è uno pseudogene che codifica una proteina mutante, coinvolta nei processi tumorali. Questi due geni, però, sono distinguibili come bersagli, da parte degli anticorpi. NANOGP8, ad esempio, è in grado di riprogrammare i fibroblasti umani. Questo pseudogene è processato e sena introni, il suo promotore risulta difficile da individuare.



I biomarcatori possono essere:

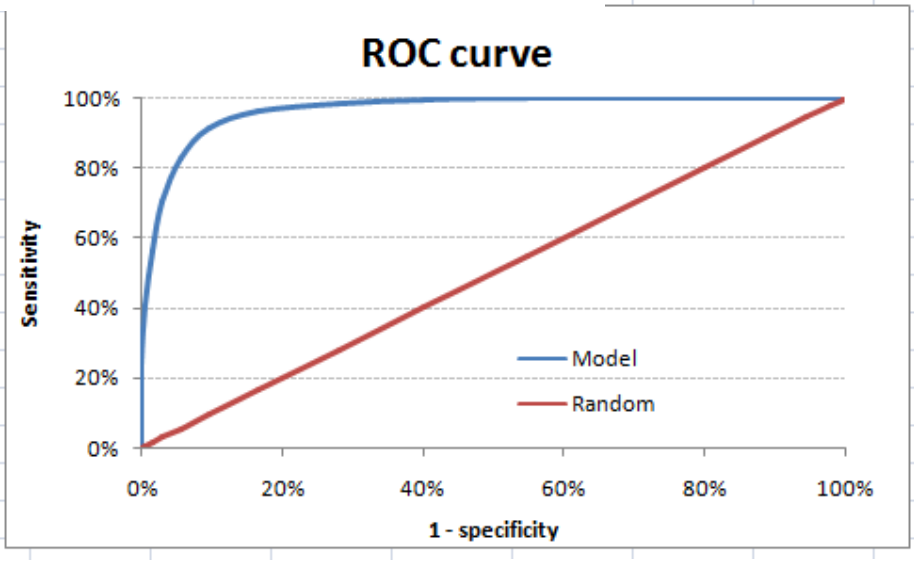
- Diagnostici, per poter affermare se una persona è affetta o meno da una malattia, come un tumore;
- Prognostici, per elaborare pronostici su quale sarà l'evoluzione del tumore;
- Predittivi, per capire quali pazienti risponderanno bene alla terapia.

Lo scopo è quello di ottenere una cura il più mirata possibile.

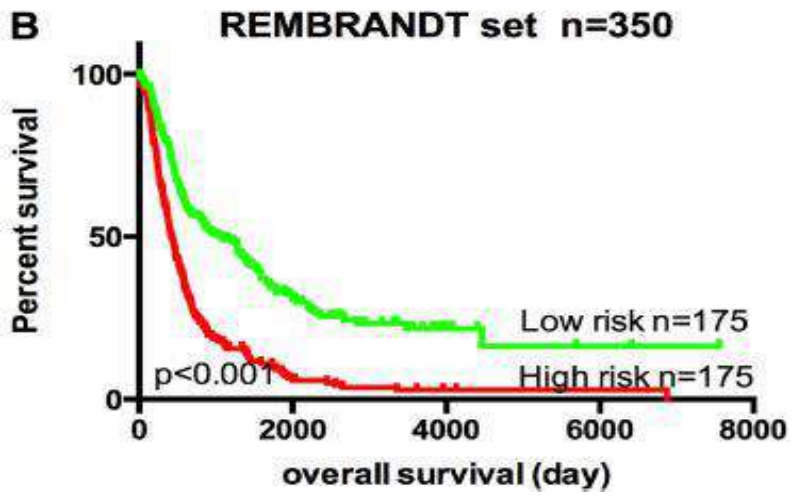


La sensibilità di un test è data dal rapporto tra i veri positivi e la somma di veri positivi più falsi negativi (il cui numero deve essere il più vicino possibile a 0, in modo tale che il rapporto totale risulti 1). La specificità, per la quale si applicano i medesimi discorsi appena enunciati, è data dal rapporto tra veri negativi e la somma di veri negativi più falsi positivi.

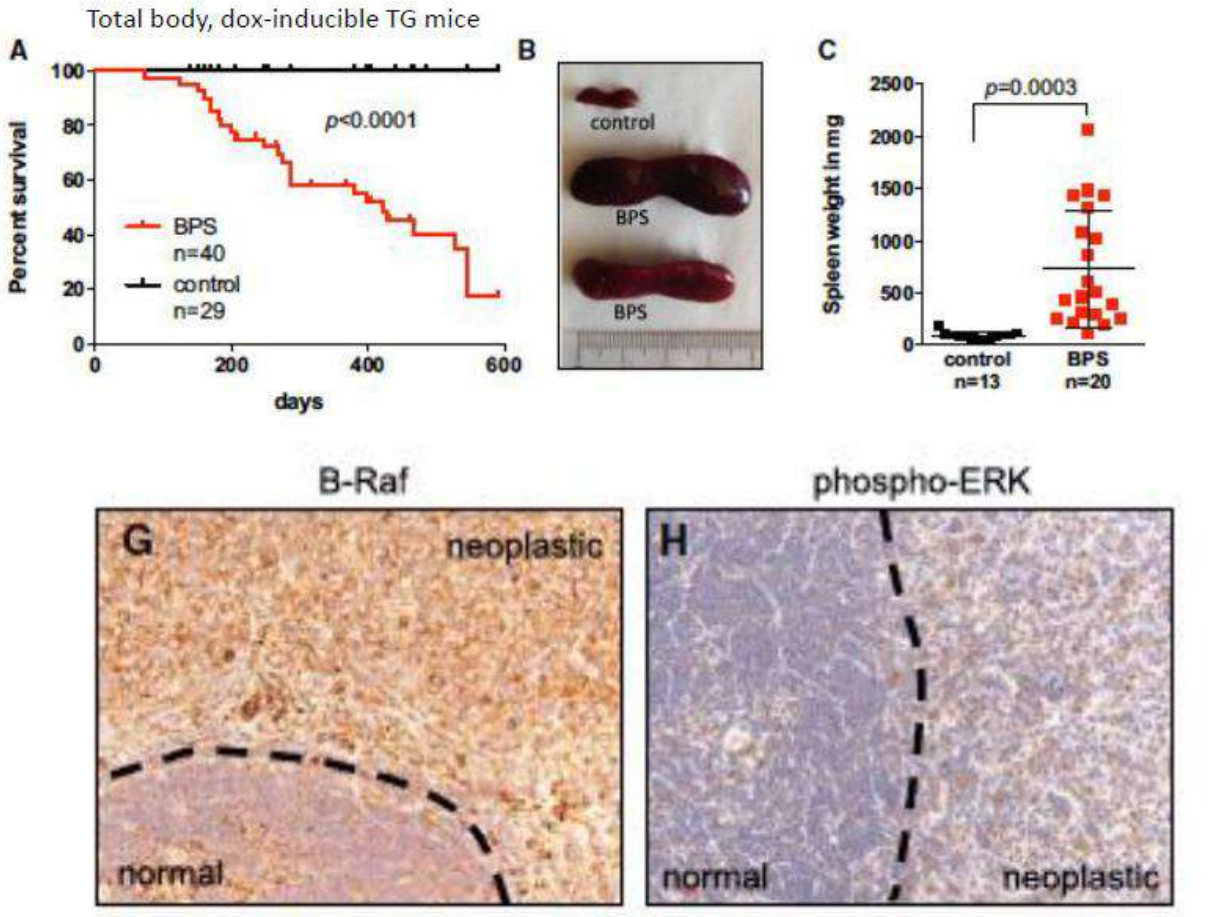
Si osserva anche la curva ROC, un grafico che illustra la performance di un classificatore binario mentre la sua soglia di variazione cambia. L'area ideale è quella al di sopra della curva, mentre l'area sotto la curva (AUC) è un indice di accuratezza del test: se vale da 0,9 a 1 la prova è eccellente, se assume valori tra 0,5 e 0,6 il test è stato inutile.



La Kaplan-Meyer, invece, considera la sopravvivenza dei pazienti: per ogni persona che è deceduta, scalo verso il basso. I pazienti affetti da glioma che presentavano un low risk, in seguito all'analisi di 6 PG (usati quindi come biomarker diagnostici), avevano un'aspettativa di vita più lunga, mentre gli high risk morivano più rapidamente.



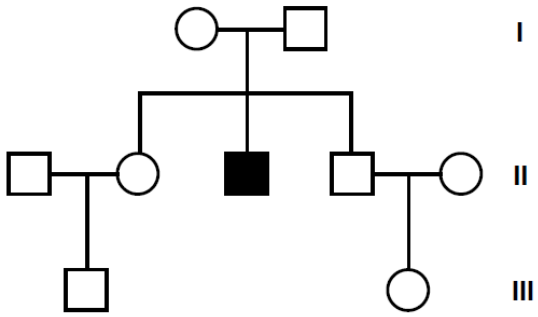
Lo pseudogene Braf-rs1, nel topo, è un ceRNA, ovvero è una spugna di miRNA (risulta favorito nel reclutamento di microRNA). Se overespresso, provoca un linfoma alla milza, causando l'amplificazione a valle di phospho-ERK.



Gli pseudogeni costituiscono sicuramente uno strumento importante nella classificazione dei tumori. Hanno relazioni estremamente complesse, con i loro geni parentali, di cui regolano l'espressione. La loro alterazione è connessa, attraverso un rapporto causa-effetto, ai tumori. I PG sono utili biomarcatori per il cancro. Dovranno essere studiati da soli, evitando le interferenze provocate dai geni parentali, e, analogamente, la scoperta non voluta di PG durante lo studio di geni codificanti deve essere severamente evitata.

ESERCIZI GENETICA PROF.SSA GEMIGNANI

Nell'albero genealogico sotto indicato è rappresentata una famiglia in cui si è verificato un caso di alcaptonuria (*a/a*, simbolo nero). Se i due cugini di primo grado (III) si sposano, con che probabilità loro figlio sarà affetto?



Si suppone che chi entra nella famiglia (II-1 e II-5) sia normale, quindi omozigote normale. Perché un soggetto sia malato i genitori devono essere eterozigoti.

Qual è la probabilità che da 2 eterozigoti nasca un figlio sano? II-2 e II-4→ $\frac{2}{3}$

La **probabilità di un evento** è il rapporto tra il numero dei casi favorevoli a quell'evento, diviso il numero totale dei casi possibili

Dall'incrocio *AAxAa* qual è la probabilità che il figlio sia eterozigote? III-1 e III-2 → $\frac{1}{2}$

Dall'incrocio *AaxAa* la probabilità che esca un figlio affetto sarà $\frac{1}{4}$

Il risultato sarà una probabilità condizionata, cioè il risultato dipende da una serie di circostanze e la probabilità si calcola facendo:

$$\frac{2}{3}\cdot\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{36}$$

La distrofia muscolare di Duchenne è associata al sesso e generalmente colpisce solo i maschi.
Coloro che ne sono affetti diventano sempre più deboli a partire dalla fanciullezza.
a. Con che probabilità una donna il cui fratello abbia la malattia di Duchenne avrà un figlio affetto?
b. Se il fratello di tua madre (tuo zio) ha la malattia di Duchenne, con che probabilità tu porti quel gene?

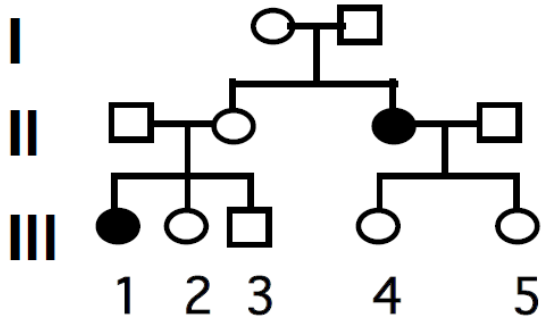
a. Dire che questa malattia è associata significa, sostanzialmente, che si tratta di una condizione patologica legata al cromosoma X. Trovandoci in questa situazione, possiamo asserire che lo zio del bambino nascituro porta l'allele recessivo (d piccolo), sul suo unico cromosoma X. Il suo genotipo sarà dunque dY. Il fatto che lui possieda un allele d comporta che questo sia stato ereditato dalla madre della coppia di fratelli: il cromosoma Y proviene dal padre e l'X dalla madre, per un individuo di sesso maschile. Il padre di entrambi i fratelli è sano, altrimenti possederemmo un'informazione relativa alla sua malattia, quindi avrà genotipo DY. La madre,

allora, appartenente alla generazione I di questo albero genealogico, è portatrice sana della distrofia, quindi sarà Dd (ci sono due cromosomi X). La figlia, sorella dell'uomo malato, avrà sicuramente un allele sano D, proveniente dal padre, mentre dalla madre può ereditare un allele malato d, con il 50 % di probabilità o un allele sano D, con altrettanta probabilità. Ora, l'uomo con il quale la donna avrà il figlio, essendo esterno alla famiglia, viene considerato sano, quindi DY. Se la moglie è sana, omozigote per l'allele D, l'incrocio sarà DY x DD e potranno nascere solo figli maschi sani. La probabilità di bambini malati maschi è uguale a 0. Se la donna è, invece, portatrice sana, l'incrocio sarà Dd x DY e c'è il 50 % di probabilità che nasca un bambino malato, dipende dall'allele donato dalla madre. Perché nasca un figlio malato, quindi, devono verificarsi contemporaneamente due condizioni: la madre deve essere portatrice sana di distrofia e deve nascere un figlio malato. Entrambe le situazioni hanno il 50 % di probabilità di verificarsi, cioè ½. Moltiplico quindi le due probabilità e ottengo ¼, che è la risposta al primo punto.

b. La risposta è sempre ¼. Deve infatti essere successo che tua madre ha ereditato da tua nonna l'allele recessivo malato e questo è successo con ½ di probabilità (vedi il ragionamento precedente). Tua madre ti avuto da un uomo che, per convenzione, si considera sano. Tua madre è perciò portatrice sana di distrofia con p = 1/2: Dd. L'incrocio è, quindi, Dd x DY. Ho quattro risultati possibili: DD, Dd (e sono entrambi individui di sesso femminile), DY e dY (bambini maschi). In ogni caso, sia che tu sia un maschio sia che tu sia una femmina, c'è ½ di probabilità (1 individuo su 2) che tu stia portando il gene della distrofia mutato. Questo ½ va, però, moltiplicato per l'1/2 di probabilità che la madre sia portatrice sana.

La fenilchetonuria (PKU) è un errore innato del metabolismo dell'aminoacido fenilalanina. La caratteristica manifestazione della PKU è un grave ritardo mentale. L'albero di tre generazioni mostrato nella figura è relativo ad una famiglia in cui è presente la malattia.

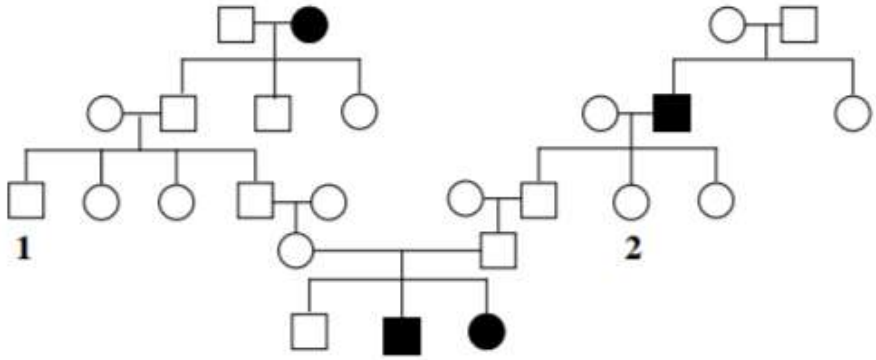
a) Qual è la modalità di ereditarietà della PKU?
b) Quali persone dell'albero sono eterozigoti per la PKU?
c) Qual è la probabilità che III 2 sia una portatrice (eterozigote)?
d) Se III 3 e III 4 si sposano, qual è la probabilità che il loro primo figlio abbia la PKU?



a) La PKU è recessiva, poiché non compare nei genitori di individui affetti, ed è autosomica, poiché i padri di donne malate sono sani.
b) Sono certamente eterozigoti tutti i genitori sani di individui affetti, quindi gli individui I-1, I-2, II-1 e II-2. Inoltre, sono certamente eterozigoti anche i figli sani di almeno un genitore malato, quindi III-4 e III-5.

c) I genitori di III-2 sono certamente eterozigoti (punto b), quindi lei sarà portatrice, essendo sana, con probabilità pari a 2/3.
d) III-3 ha probabilità 2/3 di essere portatore (vedere anche punto c), mentre III-4 è certamente eterozigote (punto b).
Se fossero entrambi eterozigoti, la probabilità di avere un figlio malato sarebbe 1/4. Applicando la regola del prodotto per eventi indipendenti, la probabilità composta è: 2/3x1x1/4=1/6.

Questo è l'albero genealogico di una famiglia in cui ricorre una rara malattia renale (nota: poiché la malattia è rara, se non ci sono evidenze del contrario assumere che gli individui esterni alla famiglia siano omozigoti sani).



a) Indicare la modalità ereditaria di questa malattia motivando la risposta.

b) Se gli individui 1 e 2 si sposassero, quale sarebbe la probabilità che il loro primo figlio fosse affetto dalla malattia?

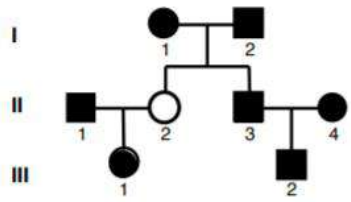
a) Gli individui affetti non hanno genitori affetti, quindi la malattia è recessiva. Inoltre non è legata né all'Y (colpisce anche le femmine) né all'X (altrimenti II-2, II-3 e IV-2 avrebbero dovuto essere malati): la malattia è autosomica.
b) L'individuo 2 è certamente eterozigote: è sana, ma ha un genitore malato. L'individuo II-1 assumiamo che sia omozigote normale, come indicato nel testo dell'esercizio. II-2 è certamente eterozigote: è sano, ma ha un genitore malato. La probabilità che l'individuo 1 sia eterozigote è dunque 1/2. La probabilità che, se 1 è eterozigote, nasca un figlio malato, è 1/4. La probabilità composta dei due eventi (regola del prodotto) è dunque: 1/2x1/4=1/8.

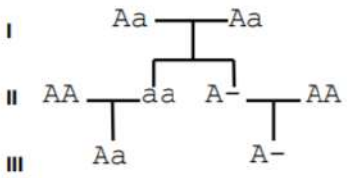
Nell'uomo la fenilchetonuria è una grave malattia metabolica dovuta ad un difetto dell'enzima che trasforma l'amminoacido fenilalanina in tirosina. Questa alterazione è determinata da un allele recessivo di un gene autosomico. Se due genitori eterozigoti hanno quattro figli:

a) qual è la probabilità che nessuno di essi avrà la fenilchetonuria?
b) qual è la probabilità che due figli ne siano affetti?

a) Siamo in presenza di un incrocio tra due individui eterozigoti, Aa x Aa. Essendo una malattia autosomica recessiva, si manifesta soltanto negli individui omozigoti recessivi. Gli omozigoti dominanti e gli eterozigoti sono sani. La progenie di questi genitori è: ¼ aa, ½ Aa, ¼ AA. Dunque, ogni volta che nasce un figlio, ci sono ¾ di probabilità che nasca sano, e ¼ che sia malato. Se i genitori hanno quattro figli e devono essere tutti sani, significa che 4 volte sono nati figli sani. Ciascun figlio ha ¾ di probabilità di nascere sano e questo è avvenuto 4 volte. L'operazione da fare è ¾ x ¾ x ¾ x ¾, ovvero (3/4)^4 = 81/256.
b) Il fatto che esattamente due figli siano malati è dato dalla seguente espressione: ¾ x ¾ (due figli sono sani) x ¼ x ¼. Non sappiamo quali due figli siano malati, però. Ci sono 6 possibili combinazioni di figli malati: il primo e il secondo, il primo e il terzo, il primo e il quarto, il secondo e il terzo, il secondo e il quarto, il terzo e il quarto. Dunque, dobbiamo moltiplicare per 6 l'espressione sovracitata: ¾ x ¾ x ¼ x ¼ x 6 = 54/256.

Il colore del pelo nelle cavie è controllato dai due alleli *A* e *a*; *A* è dominante su *a* e determina il pelo nero. Nel *pedigree* riportato sotto gli individui fenotipicamente neri sono indicati da simboli neri. In base ai dati, determinare qual è il genotipo degli individui dell'albero genealogico e stabilire la probabilità che un figlio nato dall'incrocio III-1 x III-2 abbia il pelo bianco. (Nota: se non ci sono prove del contrario, assumere che gli individui di cui sono ignoti i genitori siano omozigoti)





Dando per scontato, visto che non ci sono prove del contrario, che gli individui II-1 e II-4 siano genotipicamente *AA*, e visto che l'individuo III-1 è certamente eterozigote *Aa* (ha il pelo nero ma un genitore omozigote bianco), si ha:

a. la probabilità che II-3 sia *Aa* è 2/3;
b. se II-3 è *Aa*, allora la probabilità che III-2 sia *Aa* è 1/2;
c. se III-2 è *Aa*, allora la probabilità che da III-1 X III-2 nasca un *aa* è 1/4;

quindi, la probabilità totale (eventi indipendenti) è: 2/3 x 1/2 x 1/4 = 1/12.

In una popolazione cinese, il 99% di tutte le persone testate è Rh positivo (D).
Calcolare le frequenze dei tre genotipi DD, Dd e dd attese, assumendo che la popolazione sia all'equilibrio.

Se il 99% è Rh+, allora l'1% (ovvero 0,01 in frazioni di 1) è omozigote recessivo dd. Ponendo la frequenza di d uguale a q, si ha che la frequenza allelica di d è uguale alla radice quadrata di q, e cioè 0,1. La frequenza p di D sarà allora:
p = 1-0,1 = 0,9
Si avrà quindi: Individui DD = p2 = 0,9x0,9 = 0,81 = 81%
Individui Dd = 2pq = 2x0,9x0,1 = 0,18 = 18%

Nella popolazione umana il numero di maschi e femmine si equivale; ne discende che la frequenza del cromosoma Y è 0,25 e quella dell'X è 0,75. Tuttavia, se applicassimo ad essi la legge di Hardy-Weinberg (2pq), i maschi dovrebbero essere il 37,5% e non la metà. La popolazione umana non è all'equilibrio? Oppure non è applicabile la legge di Hardy-Weinberg? Spiegare e commentare in massimo dieci righe.

In questo caso la legge di Hardy-Weinberg non è applicabile perché di fatto gli incroci non sono casuali. Infatti, solo incroci XX x XY possono dare progenie vitale e quindi, ad esempio, non è possibile avere individui YY (che, tra l'altro, non sarebbero vitali, e questo ulteriormente impedisce l'applicazione di questa legge).

Si esamina la mobilità elettroforetica di un enzima in un campione di 300 piante della stessa popolazione, la cui variabilità dipende dal genotipo determinato da due alleli di un singolo gene, F e f. I risultati sono: 7 piante con genotipo FF, 106 con genotipo Ff e 187 con genotipo ff.
a) Quali sono le frequenze alleliche di F e f?
b) In regime di accoppiamento casuale, quali sono i numeri attesi dei tre genotipi sul campione dei 300 individui?

a) Ricordando che ogni individuo porta due copie di ciascun gene, la frequenza dell'allele F è data da: (7x2 + 106)/2x300 = 0,2.
   La frequenza dell'allele f sarà quindi: 1-0,2=0,8 ovvero: (2x187 + 106)/2x300.
b) I numeri attesi su una popolazione di 300 individui sono:
   FF=p2 x300=0,2x0,2x300=12;
   Ff=2pq x 300=2x0,2x0,8x300=96;
   ff=q2 x300=0,8x0,8x300=192.

In una popolazione si osservano le seguenti frequenze genotipiche: AA=60 Aa=20 aa=20
Calcolare le frequenze dei due alleli e le frequenze genotipiche attese se la popolazione fosse in equilibrio di Hardy-Weinberg.

La frequenza p di A è: (60x2)+20/(2x100)=0,7.
La frequenza q di a è: 1-p=0,3.
All'equilibrio si ha che: la frequenza degli individui AA è p2 , cioè: 0,7x0,7=0,49;
la frequenza degli individui Aa è 2pq, cioè: 2x0,7x0,3=0,42;
la frequenza degli individui aa è q2 , cioè: 0,3x0,3=0,09.

La fibrosi cistica è un carattere autosomico recessivo monofattoriale. In una certa popolazione in equilibrio, il numero di individui malati aa è 1/2500. Quanto valgono le frequenze genotipiche?
Nell'ipotesi di accoppiamento casuale, con che probabilità due individui eterozigoti, sul totale della popolazione, si sceglieranno come partner sessuali?
La frequenza q dell'allele recessivo è uguale alla radice quadrata di 1/2500 e vale 0,02; la frequenza p dell'allele dominante è quindi 1-0,02=0,98.
Le altre frequenze genotipiche (la frequenza di aa ovviamente è 0,0004 = 1/2500) saranno quindi:
Individui AA = p2 = 0,98x0,98 = 0,9604
Individui Aa = 2pq = 2x0,98x0,02 = 0,0392
La probabilità che due eterozigoti si scelgano è uguale alla probabilità di prendere a caso, dalla popolazione intera, i due eterozigoti stessi, quindi applicando la regola del prodotto: 2pqx2pq = 0,0015 circa.

La fenilchetonuria nell'uomo è data da una mutazione autosomica recessiva; gli individui omozigoti *a/a* presentano un quadro clinico grave e normalmente non si riproducono. Se in una popolazione all'equilibrio la frequenza degli omozigoti *a/a* è 1/10000, con che probabilità da due individui presi a caso dalla popolazione stessa nascerà un individuo omozigote *a/a*?

Poiché la popolazione è all'equilibrio, la risposta è ovviamente 1/10000.

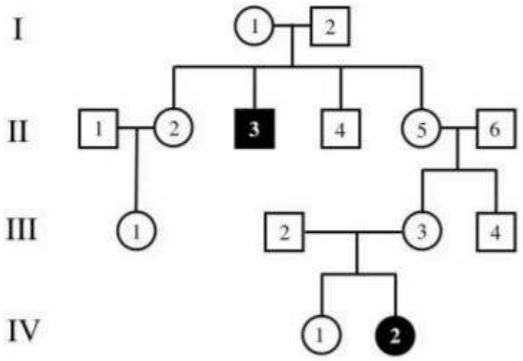
E' possibile giungere alla stessa conclusione calcolando. Affinché nasca un omozigote *a/a*, entrambi i genitori debbono essere eterozigoti (gli omozigoti *a/a* non si riproducono); da essi nascerà un omozigote *a/a* con frequenza 1/4 = 0,25. La probabilità di avere due individui entrambi eterozigoti sarà uguale al prodotto delle singole probabilità, con l'accortezza che il totale da considerare non è la popolazione globale, ma la popolazione a **meno** del numero degli omozigoti *a/a* (eterozigoti *A/a* fratto totale con fenotipo **normale**). La frequenza q dell'allele recessivo è uguale alla radice quadrata della frequenza degli omozigoti *a/a* (1/10000), quindi 0,01. La frequenza p dell'allele dominante è perciò 1-0,01=0,99. Il risultato è dato dalla seguente espressione:

$$\left(\frac{2pq}{p^2 + 2pq} \right)^2 \times \frac{1}{4}$$

ovvero:

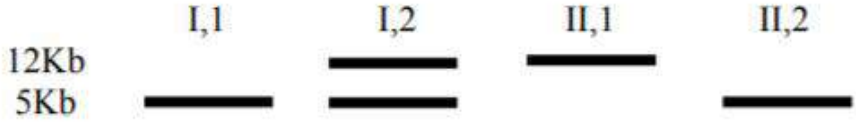
$$\left(\frac{2 \times 0{,}99 \times 0{,}01}{(0{,}99)^2 + 2 \times 0{,}99 \times 0{,}01} \right)^2 \times 0{,}25 = 0{,}0001 = 0{,}01\% = 1/10000$$

L'albero genealogico in figura rappresenta l'ereditarietà di una malattia autosomica recessiva. Assumendo che II-1 sia omozigote, calcolare con che probabilità nascerà un individuo sano dall'incrocio III-1 x III-2.



Il modo più semplice per rispondere è calcolare la probabilità che il figlio sia malato, e poi sottrarre questo risultato dall'unità. Quindi: probabilità che II-2 sia portatrice: 2/3; probabilità che III-1 sia portatrice, se lo è la madre: 1/2; probabilità che dall'incrocio richiesto nasca un individuo malato, se III-1 è portatrice, considerando che III-2 è certamente eterozigote: 1/4. Applicando la regola del prodotto, la probabilità che l'individuo richiesto sia malato è: 2/3x1/2x1/4=1/12. Per cui la probabilità che sia sano è: 1-1/12=11/12.

Il gene *d*, responsabile per la distrofia muscolare di Duchenne (recessivo X-linked) dista 4cM da un sito polimorfico per l'enzima di restrizione HpaI. Due genitori fenotipicamente normali hanno avuto un figlio maschio affetto e chiedono di effettuare una diagnosi prenatale per accertare la situazione nel nascituro che è di sesso maschile. I patterns di restrizione per i componenti la famiglia risultano essere quelli indicati nella figura sotto:



Con che probabilità il feto (II,2) è affetto? Spiegare brevemente il ragionamento seguito.

Poiché il sito polimorfico è X-linked, solo le donne possono essere eterozigoti, quindi nella tabella il padre è I,1 e la madre è I,2; inoltre il figlio malato è II,1 mentre il nascituro è II,2 (vedi testo). Il figlio II,1 malato può aver preso la malattia X-linked solo dalla madre portatrice, visto che i genitori sono entrambi fenotipicamente sani. La malattia è quindi associata strettamente all'allele da 12Kb (banda in alto). L'allele da 5Kb non è associato alla

malattia né nel padre né nella madre (entrambi sani, come già detto). Quindi il nascituro II,2, che ha l'allele da 5Kb, dovrebbe essere sano, a meno che non sia intervenuto un evento di crossing-over tra il sito polimorfico ed il gene per la distrofia. Poiché essi distano 4cM, la probabilità che nel nascituro ci sia associazione tra l'allele d e l'allele da 5Kb è il 2%, ovvero la metà del totale dei ricombinanti (l'altra metà sono i ricombinanti che associano l'allele da 12kb all'allele D del gene della distrofia).

Esiste in Drosophila una mutazione chiamata Curly (Cy) che determina arricciamento delle ali. Da un incrocio di un maschio Cy con una femmina Cy si ottennero nella popolazione della F1 207 moscerini Cy e 101 normali. Spiegare il risultato.

La mutazione Cy è dominante ed è letale in omozigosi; i due individui incrociati erano quindi entrambi eterozigoti. Questo si può dedurre dal fatto che nella progenie si ha un rapporto mendeliano (prima legge) alterato 2:1 (207 Cy rispetto a 101 normali) anziché l'atteso 3:1 (per la letalità di Cy), e dal fatto che da individui mutanti nascano individui normali (dominanza di Cy sul tipo selvatico).

La signora Rossi, fenotipicamente normale, con genitori e nonni normali, ha un bisnonno Rossi albino.
a) Qual è la probabilità che lei sia portatrice dell'albinismo?
b) Qual è la probabilità di un figlio albino se la signora Rossi si sposa con un uomo fenotipicamente sano, ma con il padre di lui albino e la madre sana e senza storia familiare di albinismo?
c) Qual è la probabilità che dal matrimonio indicato al punto precedente nascano 5 figli, tutti non albini?

Si ricordi che il carattere è autosomico; chiamiamo a l'allele per l'albinismo, e A l'allele normale. Il bisnonno Rossi (discendenza paterna) è albino, quindi omozigote a/a. Il figlio (nonno Rossi) sarà certamente A/a (è sano, ma figlio di omozigote). Il padre Rossi (sano) sarà portatore con probabilità 1/2, e avrà trasmesso il carattere alla signora Rossi con la stessa probabilità. Poiché sono eventi indipendenti, la probabilità composta che la signora Rossi sia portatrice dell'albinismo è 1/2x1/2=1/4.
b) Il marito della signora Rossi è certamente eterozigote A/a. La probabilità di avere un figlio albino dal matrimonio tra due eterozigoti è 1/4, ammesso che la signora Rossi lo sia effettivamente (punto a), quindi: 1/4x1/4=1/16.
c) La probabilità di avere un figlio sano dal matrimonio del punto precedente è 1-1/16=15/16. Di conseguenza, la probabilità di averli tutti sani (eventi indipendenti) è (15/16)^5.

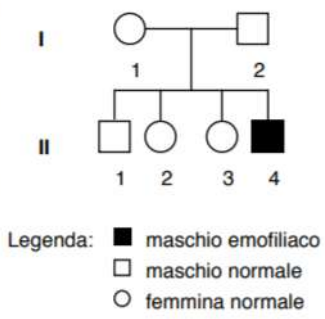
L'emocromatosi è una comune malattia autosomica recessiva che riguarda l'assorbimento del ferro. La frequenza genica è approssimativamente 0,05 (1/20) in alcune popolazioni del nord-Europa. Se la penetranza del genetipo che è causa della malattia è 0,5 nei maschi e 0,2 nelle femmine, quale è l'incidenza attesa del fenotipo della malattia nei maschi e nelle

femmine? Come si può definire, in genetica, la differenza fenotipica osservata tra uomini e donne?

Poiché la frequenza allelica è 1/20, e poiché il gene è autosomico, si deduce che la frequenza genotipica dei malati è 1/20 x 1/20 = 1/400, e questo vale sia per i maschi che per le femmine. A livello fenotipico, tuttavia, la penetranza incompleta farà sì che alcuni individui genotipicamente omozigoti recessivi non manifestino la malattia. Per sapere quanti sono si dovrà moltiplicare la frequenza ricavata sopra per il valore della penetranza, per cui: Maschi: (1/20)^2 x 0,5 = 1/800
Femmine: (1/20)^2 x 0,2 = 1/2000
Quando un carattere si esprime in modo differente nei due sessi, si parla di carattere controllato dal sesso.
NB: non è un carattere legato al sesso, poiché il gene è autosomico; non è un carattere limitato dal sesso, perché si esprime comunque in entrambi i sessi.

Un gene recessivo legato al sesso (*h*) prolunga il tempo di coagulazione del sangue, dando come risultato un'affezione nota comunemente come emofilia. Sulla base delle informazioni che si ricavano dal pedigree a lato, si risponda alle seguenti domande:

a) Se II-2 si sposa con un uomo normale, quali sono le probabilità che il suo primo figlio maschio sia affetto da emofilia?
b) Si supponga che lo sia. Quali sono le probabilità che anche il secondo lo sia?



c) Se la madre di I-1 è fenotipicamente normale, qual era il fenotipo di suo padre?

Premessa: poiché II-4 è malato, si deduce che la madre (I-1) è portatrice sana di emofilia; i genotipi dei genitori saranno quindi madre H/h e padre H/Y.
a) II-2 è portatrice con probabilità 1/2 (l'X paterno è certamente sano), trasmetterà il carattere con probabilità 1/2, il figlio sarà maschio con probabilità 1/2, quindi: 1/2x1/2x1/2=1/8.
b) II-2 è certamente eterozigote H/h, per cui trasmetterà h con probabilità 1/2 e il figlio sarà maschio con probabilità 1/2, per cui: 1/2x1/2=1/4.
c) Non è possibile stabilirlo; infatti la madre di I-1 avrebbe potuto essere portatrice H/h (e il padre quindi avrebbe potuto essere sano), oppure avrebbe potuto essere omozigote H/H e l'allele h sarebbe in questo caso venuto dal padre.

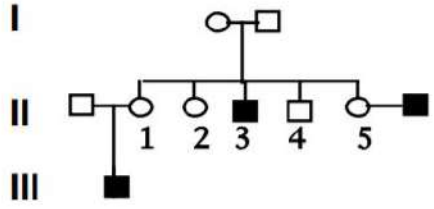
È mostrato un pannello di ibridazione di cellule somatiche che indica quali cloni (colonna 1) producono un segnale positivo di ibridazione per un segmento di DNA di un gene (colonna 2) che si intende mappare. I cromosomi umani presenti in ciascun clone sono indicati con un + nelle righe corrispondenti. In base al *pattern* di ibridazione del pannello, su quale cromosoma è collocato il segmento di DNA?

Clone	DNA segment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
1	–	–	–	–	+	–	+	+	+	+	–	+	+	–	–	+	–	+	–	+	–	–	–	+	–
2	–	+	+	+	–	+	–	+	–	–	+	–	–	–	+	+	+	–	–	+	–	+	+	+	–
3	+	–	–	–	–	+	+	+	+	+	–	–	–	+	–	+	–	+	–	+	+	–	+	+	–
4	+	+	–	+	–	–	–	+	–	+	–	+	+	–	–	–	+	–	+	–	+	+	–	–	–
5	+	–	+	+	+	+	+	–	–	–	+	–	–	+	–	+	–	+	–	–	+	+	–	+	+
6	–	+	–	+	–	–	+	–	–	–	+	+	–	+	+	–	+	–	+	–	–	–	+	+	–
7	+	–	+	–	+	–	+	+	+	+	–	+	–	–	–	+	+	–	+	–	+	+	–	+	–
8	–	–	+	+	+	–	–	+	–	–	+	+	+	–	+	–	+	+	–	+	–	–	–	–	–

Il segmento di DNA cercato è collocato sul cromosoma 20.

È stato eseguito uno screening su una popolazione per individuare mutazioni nel gene della fibrosi cistica (autosomica recessiva) usando campioni di DNA. Il test mostra che 1 individuo su 100 in questa popolazione è portatore eterozigote della mutazione della fibrosi cistica (CF).
a) Qual è la proporzione di individui affetti che ci si aspetta nella popolazione?
b) In base alla domanda precedente, se un portatore noto della CF si accoppia con qualcuno nella popolazione in generale, con quale probabilità la coppia avrà un figlio affetto da fibrosi cistica?
c) Si immagini che il portatore noto di CF della domanda precedente si accoppi con una cugina di primo grado. Con quale probabilità la coppia avrà un figlio affetto da fibrosi cistica?
a) Poiché 1/100 individui è portatore, la frequenza q dell'allele è 1/200. Di conseguenza, la probabilità di avere omozigosi è q2 =1/200x1/200= 1/40.000.
b) Probabilità che si prenda a caso nella popolazione un individuo portatore: 1/100. Probabilità che da due portatori nasca un malato: 1/4. Probabilità composta: 1/100x1/4=1/400.
c) Poiché i cugini primi hanno probabilità 1/8 di avere le stesse mutazioni, e poiché da due eterozigoti nascono figli malati con probabilità 1/4, la probabilità composta sarà 1/8x1/4=1/32.

Gli individui, indicati con simboli pieni nel seguente albero genealogico di una famiglia umana, hanno una forma comune di daltonismo, determinata da un allele recessivo legato al cromosoma X.

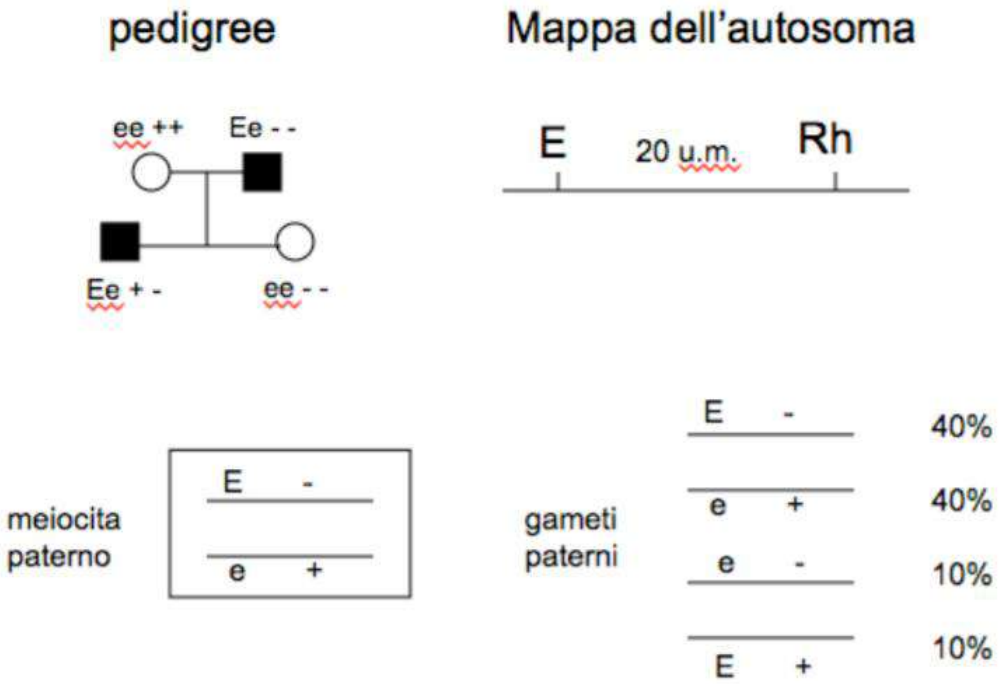


a. Se la donna identificata come II-1 ha altri due figli, qual è la probabilità che entrambi abbiano una normale visione del colore?
b. Qual è la probabilità che il primo figlio di II-5 abbia questa forma di daltonismo?

La donna II-1 è certamente portatrice (vedi III-1). Essendo il padre di III-1 sano, non è possibile che nascano figlie femmine daltoniche. Quindi i figli, per essere malati, dovranno ciascuno ricevere l'allele del daltonismo dalla madre eterozigote (probabilità: 1/2) ed essere maschi (probabilità: 1/2). Probabilità composta per eventi indipendenti: 1/2x1/2=1/4 (maschio daltonico), quindi ognuno sarà sano con probabilità 3/4, e perciò saranno entrambi sani con probabilità 3/4x3/4=9/16.
La madre di II-5 è portatrice (vedi II-3) ma suo padre (I-2) è sano, quindi II-5 sarà portatrice con probabilità 1/2. Il marito di II-5 è daltonico, quindi se II-5 è portatrice possono nascere anche femmine daltoniche.
Se la madre fosse portatrice (1/2) e trasmettesse il carattere (1/2) tutti i figli, indipendentemente dal sesso, sarebbero daltonici. Quindi la presenza della malattia dipenderà solo dai due eventi appena descritti. La probabilità composta per eventi indipendenti è quindi: 1/2x1/2=1/4.

I geni per l'ellissocitosi e per l'antigene Rh degli eritrociti sono associati su uno degli autosomi e distano approssimativamente 20 unità di mappa. Un uomo con ellissocitosi (malattia dominante), la cui madre ha eritrociti di forma normale ed un genotipo omozigote Rh+ ed il cui padre è Rh- ed eterozigote per ellissocitosi, si sposa con una donna normale Rh-.
A) Quali sono le probabilità che il loro primo figlio sia Rh- ed affetto da ellissocitosi
B) Se esso è invece Rh+, quali sono le probabilità che sia anche affetto da ellissocitosi?

Chiamiamo E l'allele dominante del gene dell'ellissocitosi ed e il corrispondente gene recessivo (normale). Il testo dell'esercizio ci permette di determinare il genotipo dell'uomo e della partner (vedi figura). Inoltre, dal momento che ci viene fornita la distanza di mappa tra i geni E ed Rh siamo in grado di calcolarci la frazione di gameti che portano la combinazione allelica parentale o ricombinante nella meiosi dell'uomo che è un eterozigote.



A) perché il figlio sia Rh- ed affetto da ellissocitosi bisogna che erediti dal padre l'allele dominante E insieme all'allele Rh- ovvero una combinazione parentale degli alleli. La probabilità che questo avvenga è pari al 40% (la madre è omozigote per entrambi i geni e trasmetterà ai figli soltanto combinazioni recessive dei due geni.

B) Se il figlio è Rh+ può avere ricevuto o la combinazione parentale e + o quella ricombinante E +.
Dal momento che soltanto il 20% dei gameti Rh+ (nota bene non del totale dei gameti) sono anche E, 20% sarà la probabilità che un figlio Rh+ abbia l'ellissocitosi.

Una donna di 30 anni, sana, il cui fratello è affetto da galattosemia (malattia autosomica recessiva rara), si presenta al consultorio e vi chiede qual è la probabilità di essere portatrice di questa malattia. Se sposa un uomo sano, non parente, vi attendete che i figli di questa coppia saranno malati?

Il fratello della donna ha chiaramente genotipo aa, quindi può essere nato soltanto da due genitori eterozigoti, Aa x Aa. 4 sono gli esiti possibili: AA, Aa, Aa, aa. Sappiamo che la donna è sana, quindi dal calcolo della probabilità dobbiamo escludere il caso aa. Rimangono quindi due scenari possibili: AA (omozigote sana) e Aa (effettivamente portatrice). Le due situazioni non sono però equiprobabili, perché è più facile che sia portatrice. In particolare, lo è in 2 casi su 3, quindi con 2/3 di probabilità.
L'uomo che entra nella famiglia è sano, non parente, quindi possiamo considerarlo AA. Ci sono due incroci possibili: AA x AA (mamma sana) e AA x Aa (mamma portatrice). Nel primo caso la progenie sarebbe solo AA, quindi costituita da individui sani. Nel secondo caso la progenie sarebbe formata da individui sani o portatori sani. In entrambi i casi non è attesa la nascita di figli malati.

Un uomo affetto da una rara malattia autosomica recessiva a penetranza incompleta (80%) sposa una propria parente che sappiamo essere eterozigote per quel carattere. Che previsione fate circa la comparsa della malattia nella prole?

L'uomo ha genotipo aa, mentre la moglie è sicuramente Aa. Abbiamo un incrocio Aa x aa. Si avranno individui portatori con probabilità ½ e individui malati con altrettanta probabilità. La penetranza è all'80%, quindi significa che su 10 omozigoti, che presentano lo stesso genotipo (aa), 8 soltanto mostrano il fenotipo atteso, in questo caso, quello malato. 2 su 10, nonostante siano omozigoti recessivi, non sono malati. Dobbiamo perciò moltiplicare la probabilità di essere omozigoti malati (1/2) per 8/10. Il risultato è 8/20, cioè 2/5. La probabilità di avere figli malati è 0,4 leggermente inferiore al valore atteso di 0,5 per effetto della penetranza incompleta.

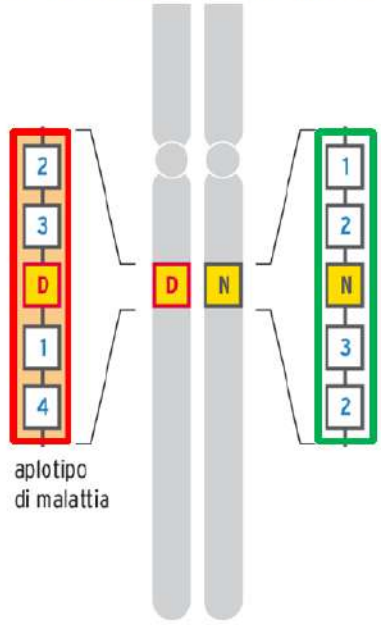
Una donna è portatrice di una rara patologia X-linked recessiva che presenta penetranza incompleta (70%). Che previsioni fate circa la presenza della malattia nella progenie?

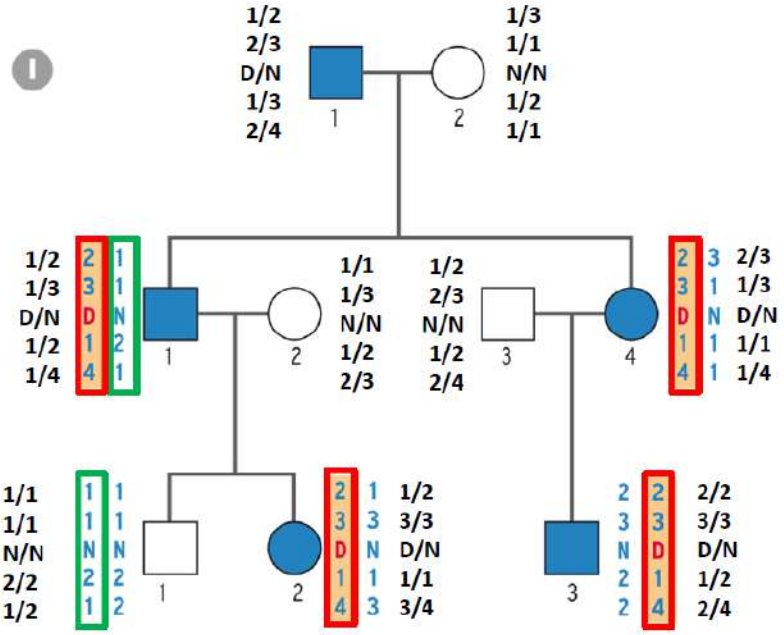
L'incrocio è tra una donna X^*X e un uomo, convenzionalmente sano, XY. Da questo incrocio otteniamo quattro individui: X^*X (femmina portatrice sana, dunque non malata), XX (femmina sana), XY (maschio sano), ciascuno con probabilità ¼. Le femmine non saranno mai malate, mentre ½ dei maschi sì. In realtà, dal momento che la penetranza è al 70%, questa probabilità sarà leggermente inferiore. Infatti, non tutti gli individui X^*Y saranno malati, ma solo 7 su 10. Dobbiamo perciò moltiplicare ¼ (probabilità di avere un bambino maschio malato) per 7/10 e otteniamo 7/40, cioè 0,175.

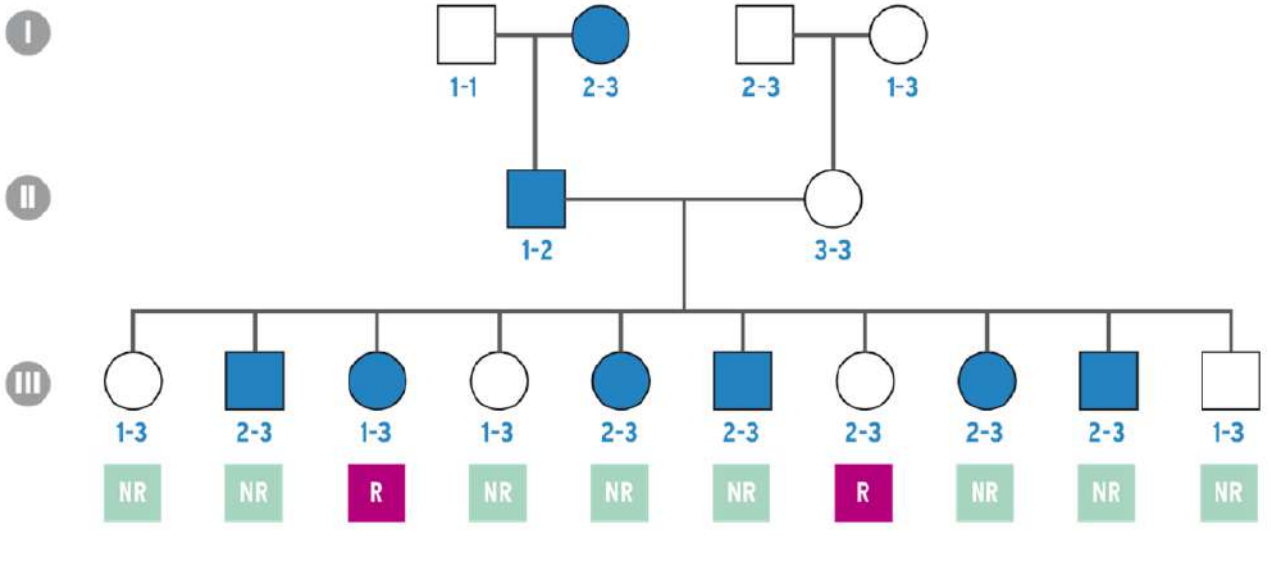
CAP. 8 PAGG. 236-237 E PAG. 239 LIBRO STRACHAN

a) Distinguere i due crom del genitore;
b) Determinare la fase;
c) Determinare quale cromosoma sia stato trasmesso alla prole.

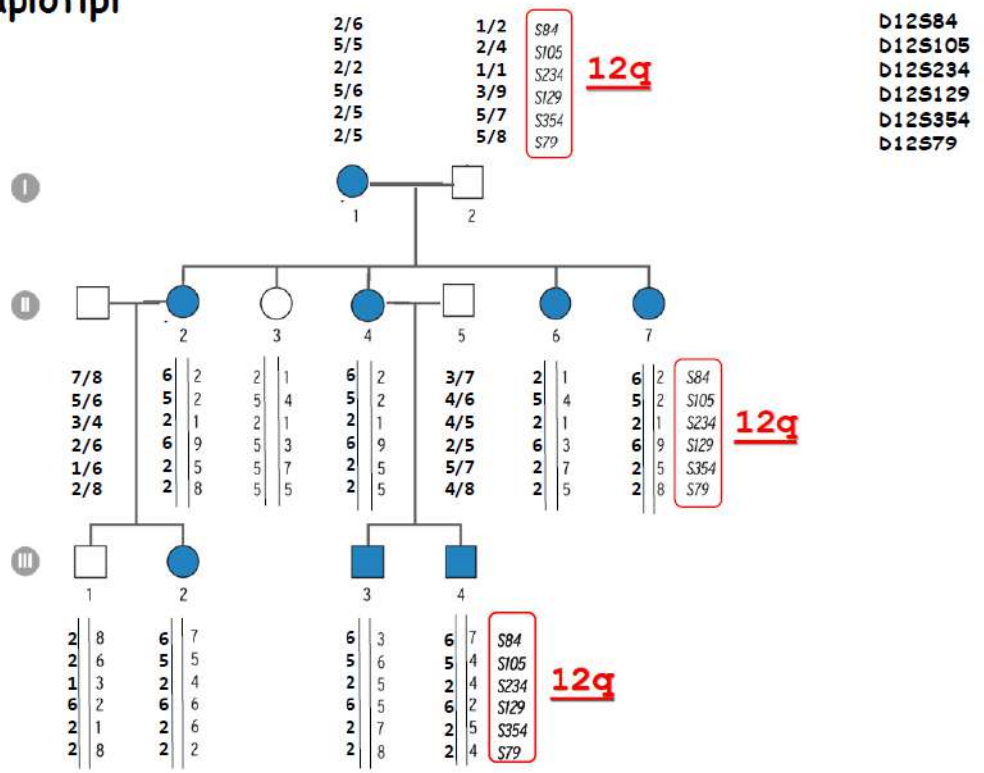
- Aplotipo malattia



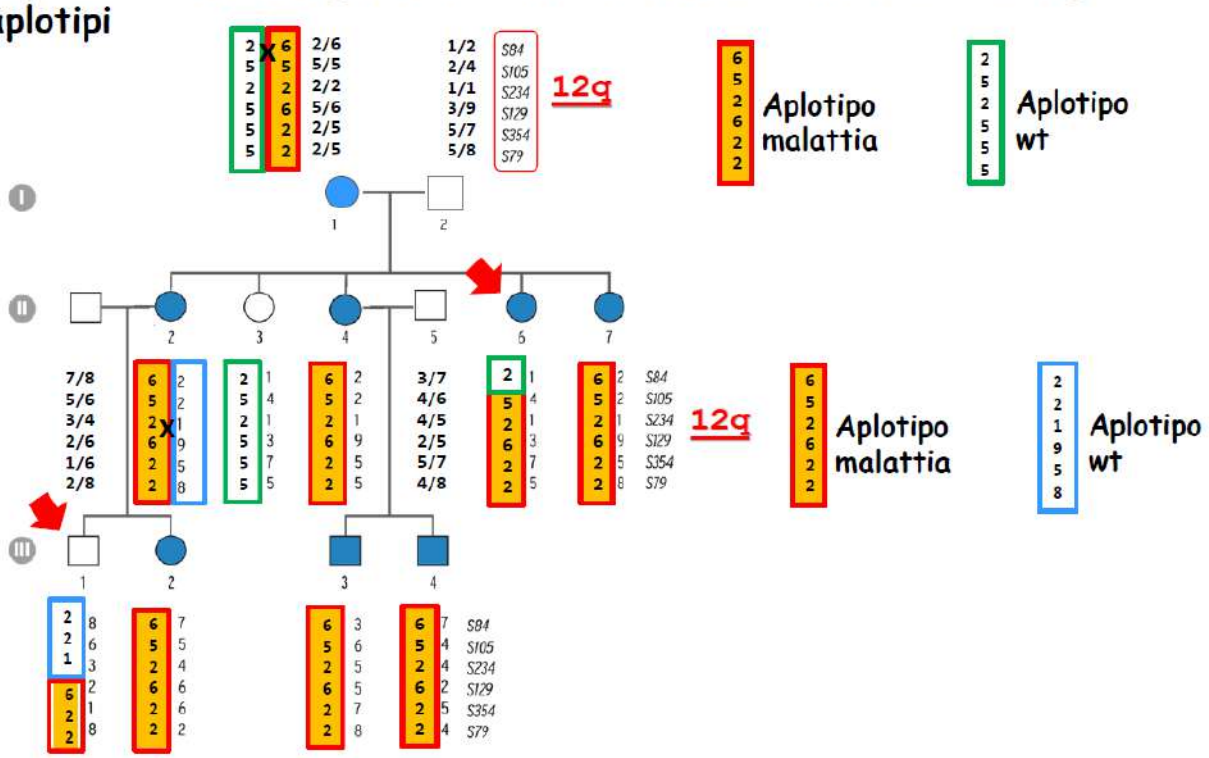


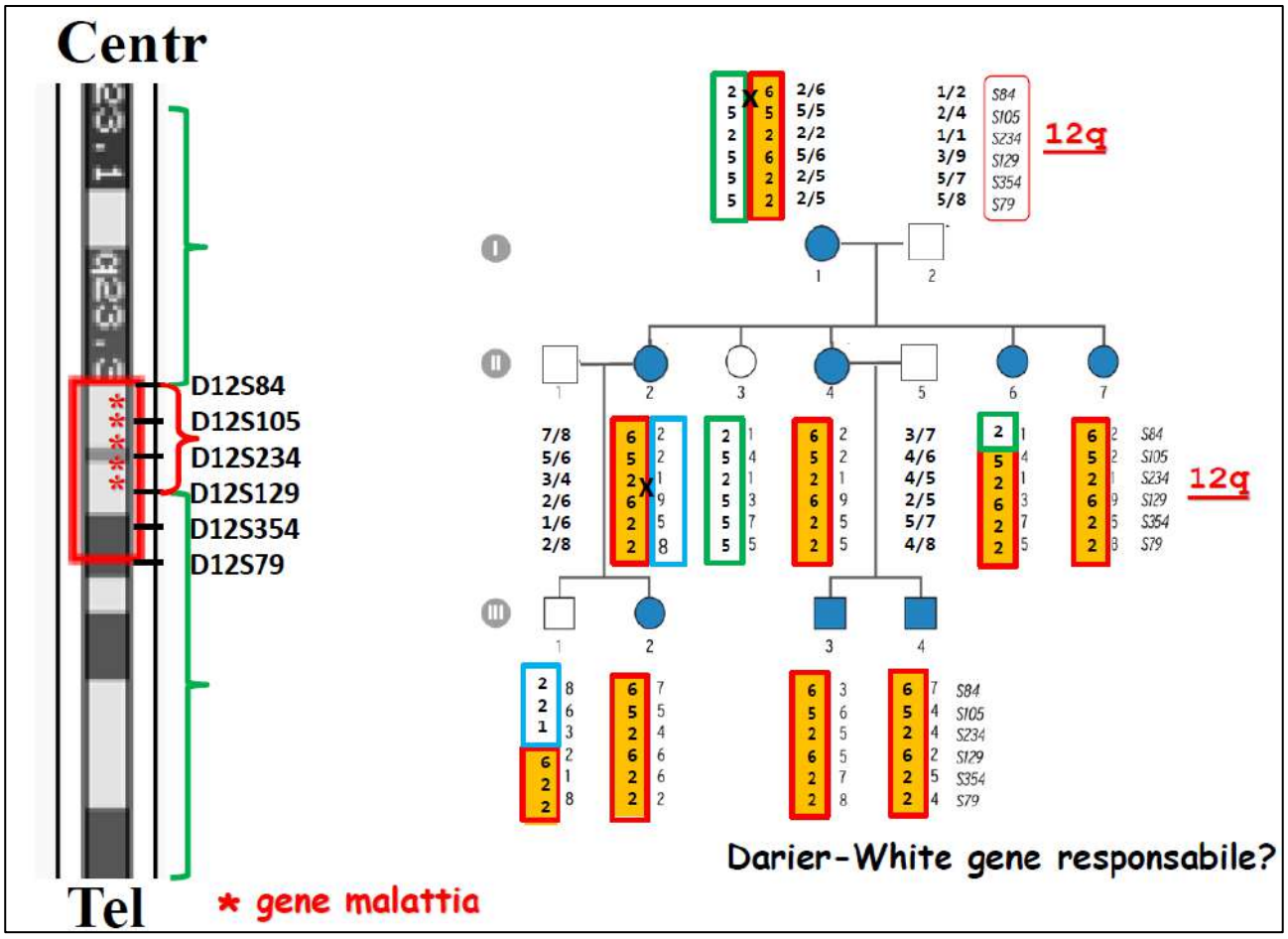


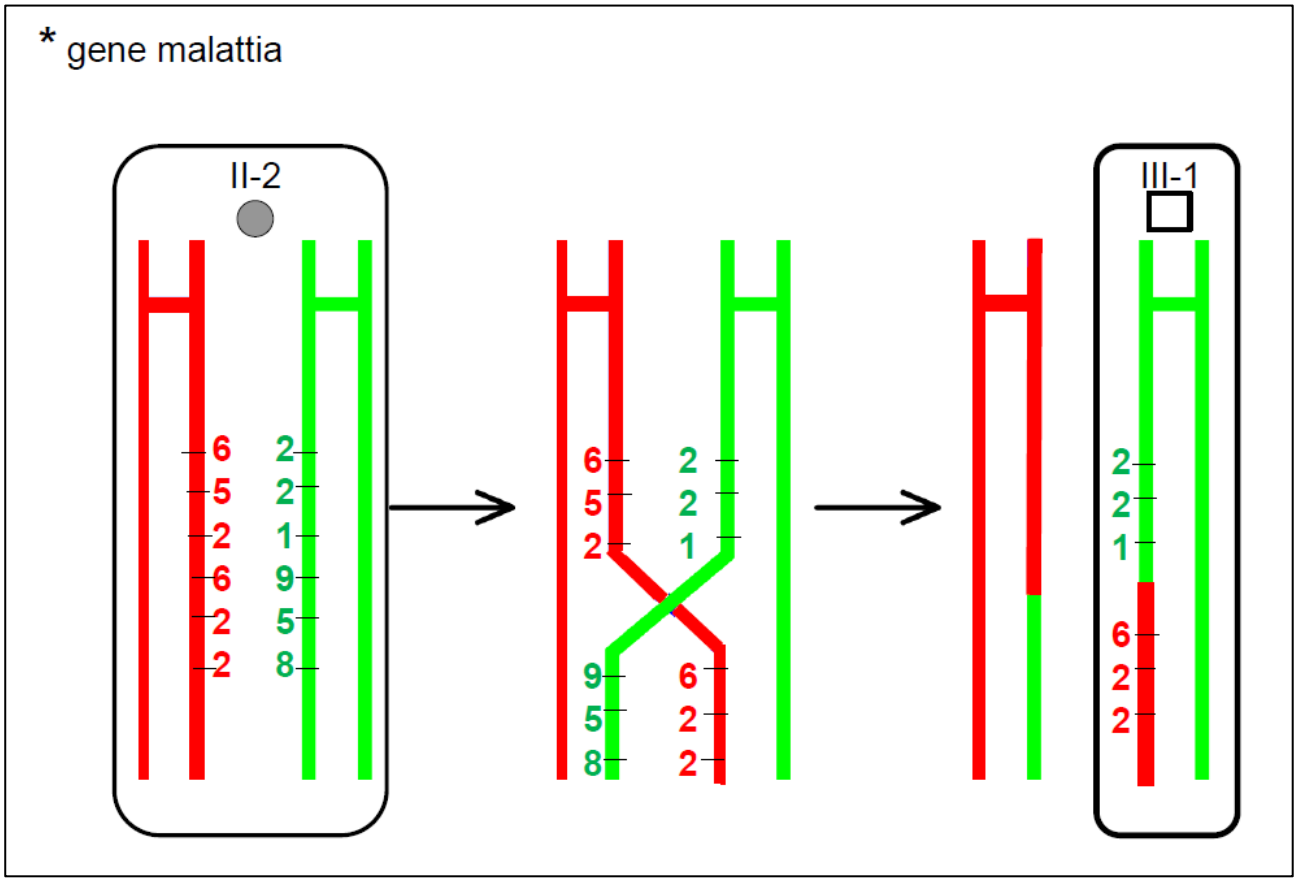
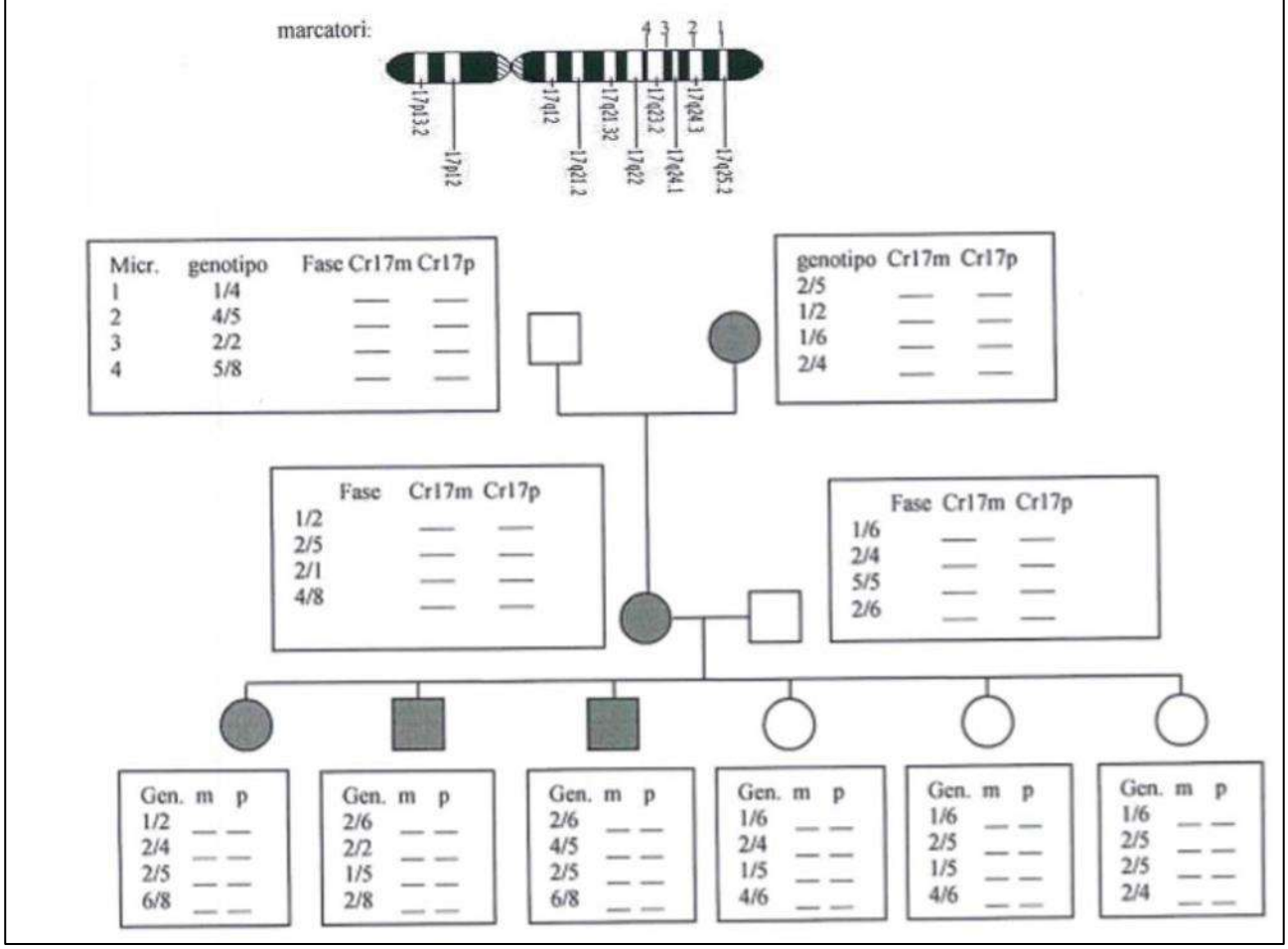
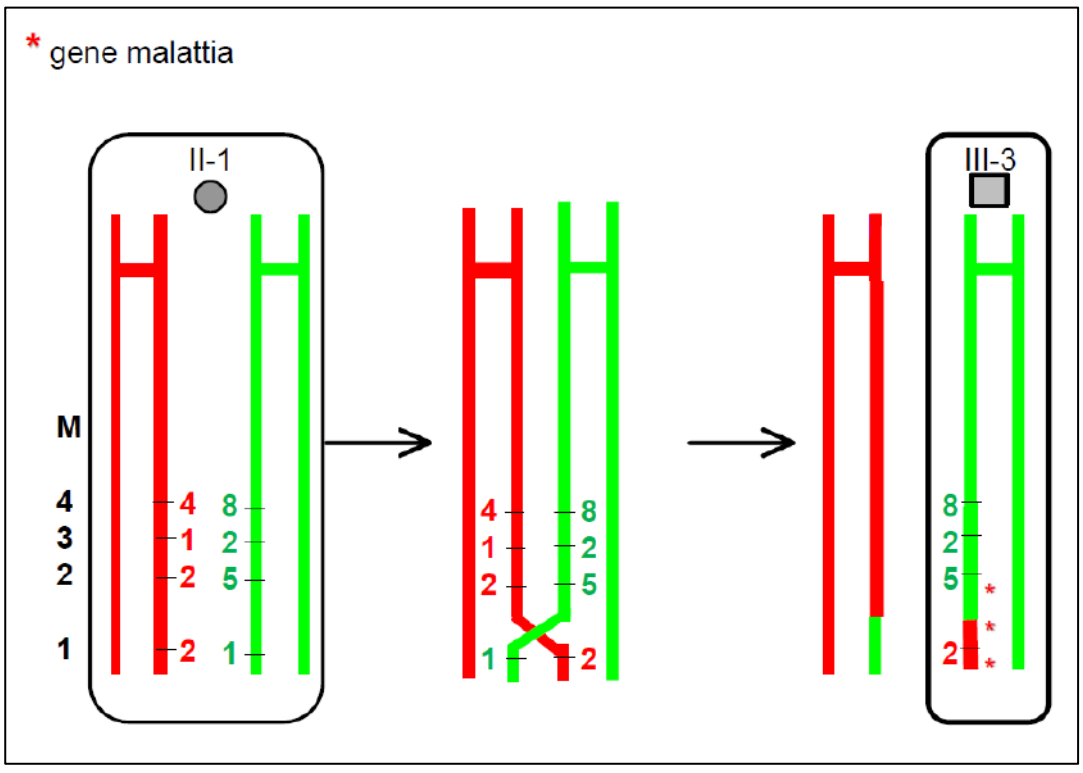
- Definizione della regione candidata minima tramite analisi degli aplotipi



- Definizione della regione candidata minima tramite analisi degli aplotipi




Centr
D12S84
D12S105
D12S234
D12S129
D12S354
D12S79
Tel
* gene malattia
12q
12q
Darier-White gene responsabile?


* gene malattia
II-2
III-1

Esercizio: Indicare la regione candidata del cromosoma 17 che potrebbe contenere il gene responsabile di un carattere autosomico dominante.